

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
DE AUTOMAÇÃO E SISTEMAS**

Felipe Sanches Gurgel

**ESTIMAÇÃO DE TEMPOS DE CHEGADA DE ÔNIBUS
URBANO UTILIZANDO FILTROS DE KALMAN**

Florianópolis

2012

Felipe Sanches Gurgel

**ESTIMAÇÃO DE TEMPOS DE CHEGADA DE ÔNIBUS
URBANO UTILIZANDO FILTROS DE KALMAN**

Dissertação submetida ao Programa
de Pós-Graduação em Engenharia de
Automação e Sistemas para a obten-
ção do Grau de Mestre em Engenharia
de Automação e Sistemas.

Orientador: Prof. Werner Kraus Ju-
nior, Dr.

Coorientador: Prof. Ubirajara Franco
Moreno, Dr.

Florianópolis

2012

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Gurgel, Felipe Sanches

Estimação de tempos de chegada de ônibus urbano
utilizando filtros de Kalman [dissertação] / Felipe Sanches
Gurgel ; orientador, Werner Kraus Junior ; co-orientador,
Ubirajara Franco Moreno. - Florianópolis, SC, 2012.

118 p. ; 21cm

Dissertação (mestrado) - Universidade Federal de Santa
Catarina, Centro Tecnológico. Programa de Pós-Graduação em
Engenharia de Automação e Sistemas.

Inclui referências

1. Engenharia de Automação e Sistemas. 2. Predição de
tempos de chegada de ônibus. 3. Filtros de Kalman. 4.
Sistemas inteligentes de transporte. 5. Teste de Kolmogorov-
Smirnov. I. Kraus Junior, Werner. II. Moreno, Ubirajara
Franco. III. Universidade Federal de Santa Catarina.
Programa de Pós-Graduação em Engenharia de Automação e
Sistemas. IV. Título.

Felipe Sanches Gurgel

ESTIMAÇÃO DE TEMPOS DE CHEGADA DE ÔNIBUS URBANO UTILIZANDO FILTROS DE KALMAN

Esta Dissertação foi julgada aprovada para a obtenção do Título de “Mestre em Engenharia de Automação e Sistemas”, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Automação e Sistemas.

Florianópolis, 27 de agosto 2012.

Prof. Jomi Fred Hübner, Dr.
Coordenador do Curso

Banca Examinadora:

Prof. Werner Kraus Junior, Dr.
Presidente

Prof. Werner Kraus Junior, Dr.
Orientador

Prof. Ubirajara Franco Moreno, Dr.
Coorientador

Prof. Luiz Alberto Koehler, Dr.

Prof. Rodrigo Castelan Carlson, Dr.

Prof. Eugênio de Bona Castelan Neto, Dr.

A Deus e à minha família.

AGRADECIMENTOS

Em especial, agradeço a Deus, que me deu esta oportunidade e sempre teve mais fé em mim do que eu mesmo.

À toda minha família, em especial meus pais, Mari e Eduardo, e minha noiva Lia, pelos exemplos de humildade, força de vontade, seriedade e dedicação, nos quais tentei me espelhar durante todo o desenvolvimento deste trabalho. Também agradeço a Rodrigo, Sandra, meus irmãos Victor e Octávio, meus avós Neusa e Laércio, por todo o incentivo e ajuda em todos os momentos.

Ao professor Marcos Vallim, que foi quem iniciou em mim o sonho de uma carreira acadêmica.

Aos meus orientadores Werner e Ubirajara, pelo suporte e auxílio em minhas limitações e dificuldades encontradas durante todo o desenvolvimento deste trabalho.

Aos amigos Thiago Bergamasco, Paulo Franco, Luciano Barreto, Gustavo Toscano, Vitor Moraes, Rafael Maestrelli, Thiago Lima, Vanderlei Weber, Günter Loch, Cassiano Tartari, Rodrigo Lange, Fábio Mantelli, Tauã Brandão e Adriano Curvo, pela presença e suporte.

Ao financiamento fornecido pela bolsa CAPES/REUNI, o qual possibilitou dedicação exclusiva à pesquisa e desenvolvimento deste trabalho.

Nunca deixe que lhe digam que não vale a pena acreditar no sonho que se tem, ou que seus planos nunca vão dar certo, ou que você nunca vai ser alguém.

Renato Russo

RESUMO

A fim de diminuir congestionamentos, poluição do ar, consumo de combustível, entre outros, tem-se buscado constantemente o desenvolvimento e modernização de sistemas de transporte público, pois sistemas mais eficientes, confortáveis e convenientes atraem um maior número de pessoas. A aplicação de linhas de ônibus em ambientes urbanos, tem sido um dos modos de transporte público mais utilizados. Para operação eficiente destas linhas, é importante conhecer a posição do veículo em tempo real, possibilitando o controle dos instantes de partida e a implantação de sistemas de informação sobre chegadas futuras, melhorando a percepção de qualidade do serviço prestado. A predição dos tempos de chegada do ônibus depende de uma série de fatores (por exemplo, atrasos em interseções sinalizadas, número de passageiros em pontos de parada, etc.). Estes fatores aumentam significativamente o nível das incertezas associadas ao processo e à medição. Este trabalho apresenta um algoritmo para predição dos tempos de chegada de ônibus urbano em pontos de parada, utilizando a abordagem de filtro de Kalman com análise de dados históricos. Fatores que aumentam o nível das incertezas associadas ao processo e à medição são considerados como propriedades estocásticas das perturbações do processo. A geração de dados de medição é realizada através de dois cenários distintos desenvolvidos e simulados no *software* de simulações microscópicas Aimsun 6.1. O teste de ajustamento de Kolmogorov-Smirnov é aplicado para análise das distribuições estatísticas destes dados. Os parâmetros necessários para configuração do filtro de Kalman são obtidos a partir de dados históricos através de dois métodos de análise estatística propostos: o método de análise longitudinal e o método de análise transversal. O filtro de Kalman é utilizado para estimação de dois estados do veículo, sua posição e sua velocidade. Por fim, é proposto um algoritmo que utiliza as estimações oriundas do filtro de Kalman para realizar a predição dos tempos de chegada do ônibus em pontos de parada.

Palavras-chave: Predição de tempos de chegada de ônibus. Filtro de Kalman. Sistemas inteligentes de transporte. Teste de Kolmogorov-Smirnov.

ABSTRACT

In order to reduce traffic congestions, air pollution, fuel consumption, and others, it has been constantly sought the development and modernization of public transportation systems, because more efficient, comfortable, and convenient systems attract more people. The application of bus lines in urban environments has been one of the public transportation modes most used. However, for efficient operation of bus lines, its important to know the vehicle position in real-time, enabling the control of departure times and implantation of information systems about future arrivals. The prediction of bus arrival time depends on a number of factors (e.g., delays at signalized intersections, number of passengers at bus stops, etc.). These factors increase significantly the level of uncertainties associated to process and measurement. This work presents an algorithm for prediction of bus arrival times at bus stops using Kalman filter with historical data analysis approach. Factors that increase the level of uncertainties associated to process and measurement are considerate stochastic properties of process disturbance. The generating of measurement data is performed by two different scenarios developed and simulated on the microscopic simulation software Aimsun 6.1. The Kolmogorov-Smirnov goodness of fit test is applied for analysis of the statistical distributions of these data. The parameters required for configuration of the Kalman filter are obtained from historical data through two proposed methods of statistical analysis: the method of longitudinal analysis, and the method of transversal analysis. The Kalman filter is used for estimation of two vehicle states, its position and its velocity. Finally, an algorithm is proposed that uses the estimates given by the Kalman filter to perform the prediction of bus arrival times at bus stops.

Keywords: Prediction of bus arrival times. Kalman filter. Intelligent transportation systems. Kolmogorov-Smirnov test.

LISTA DE FIGURAS

Figura 1 A constelação de satélites para GPS. FONTE: (JOHN et al., 2003)	13
Figura 2 Estatística D do teste de Kolmogorov-Smirnov. Baseado em Press et al. (2002).	17
Figura 3 Método gráfico para aplicação do teste de Kolmogorov-Smirnov. Baseado em Massey (1951).	19
Figura 4 $S_n(x)$ e $F_X(x)$ do exemplo. Baseado em Soong (2004)..	21
Figura 5 O ciclo do filtro de Kalman. A atualização de tempo projeta a próxima estimativa do estado atual. A atualização de medição ajusta a estimativa projetada através de medição real naquele tempo. Baseado em Welch e Bishop (2001).	26
Figura 6 Operação do filtro de Kalman. Baseado em Welch e Bishop (2001).	27
Figura 7 Representação do método de análise longitudinal para aplicação em dados de posição.	31
Figura 8 Subtrações e média resultantes da aplicação do método de análise longitudinal de dados de posição.	31
Figura 9 Representação gráfica do método de análise transversal de dados de posição.	33
Figura 10 Variâncias a cada t para dados de posição do ônibus. ..	33
Figura 11 Cenário de simulação com 16 pontos de parada.	42
Figura 12 Resultado do teste de Kolmogorov-Smirnov para os dados do primeiro cenário. Comparação entre $S_n(v)$ e $F_V(v)$	44
Figura 13 Estimação da posição do ônibus até a 15ª amostra.	47
Figura 14 Erros de estimação das posições do ônibus ao longo do trajeto em relação às posições observadas sem ruído.	47
Figura 15 Velocidades estimadas para os dados de medição do primeiro cenário.	48
Figura 16 Tempo efetivo Γ_{b_8} , tempo hipotético Ω_{b_8} e tempo estimado Υ_{b_8} de chegada do veículo ao ponto de parada.	51
Figura 17 Erro de estimação do tempo de chegada do ônibus ao ponto de parada b_8 . (a) Erro de estimação; (b) Erro absoluto de estimação.	51
Figura 18 Tempo efetivo $\Gamma_{b_{14}}$, tempo hipotético $\Omega_{b_{14}}$ e tempo esti-	

mado $\Upsilon_{b_{14}}$ de chegada do veículo ao ponto de parada.	52
Figura 19 Erro de estimação do tempo de chegada do ônibus ao ponto de parada b_{14} . (a) Erro de estimação; (b) Erro absoluto de estimação.	52
Figura 20 Erro de estimação dos tempos de chegada do ônibus aos pontos de parada $b_3, b_4, b_5, \dots, b_{16}$ em relação ao ponto b_2	53
Figura 21 Resultado do teste de Kolmogorov-Smirnov para os dados do vetor E . Comparação entre $S_n(\zeta)$ e $F_E(\zeta)$	54
Figura 22 Cenário de simulação com 44 pontos de parada.	55
Figura 23 Resultado do teste de Kolmogorov-Smirnov para os dados do segundo cenário.	57
Figura 24 Estimação de posição do ônibus entre as amostras 70 e 80.	60
Figura 25 Erro de estimação da posição do ônibus ao longo do trajeto em relação às posições observadas sem ruído.	61
Figura 26 Velocidades estimadas para os dados de medição do segundo cenário.	61
Figura 27 Tempo efetivo Γ_{b_5} , tempo hipotético Ω_{b_5} e tempo estimado Υ_{b_5} de chegada do veículo ao ponto de parada.	64
Figura 28 Erro de estimação do tempo de chegada do ônibus ao ponto de parada b_5 . (a) Erro de estimação; (b) Erro absoluto de estimação.	64
Figura 29 Tempo efetivo $\Gamma_{b_{25}}$, tempo hipotético $\Omega_{b_{25}}$ e tempo estimado $\Upsilon_{b_{25}}$ de chegada do veículo ao ponto de parada.	65
Figura 30 Erro de estimação do tempo de chegada do ônibus ao ponto de parada b_{25} . (a) Erro de estimação; (b) Erro absoluto de estimação.	65
Figura 31 Tempo efetivo $\Gamma_{b_{40}}$, tempo hipotético $\Omega_{b_{40}}$ e tempo estimado $\Upsilon_{b_{40}}$ de chegada do veículo ao ponto de parada.	66
Figura 32 Erro de estimação do tempo de chegada do ônibus ao ponto de parada b_{40} . (a) Erro de estimação; (b) Erro absoluto de estimação.	66
Figura 33 Curva em forma de sino característica de uma distribuição normal. Baseado em Trosset (2009).	88

LISTA DE TABELAS

Tabela 1	Comparação de técnicas para predição de tempos de chegada de ônibus urbano.....	10
Tabela 2	Valores críticos $d_\alpha(n)$ para diferença absoluta máxima entre duas funções de distribuição acumulada.....	18
Tabela 3	Valores da distribuição teórica $F_V(v)$ para cada amostra v	44
Tabela 4	Erro percentual ξ_{b_i} para o erro médio ϵ_{b_i} da estimativa do tempo de chegada em relação ao tempo efetivo Γ_{b_i}	50
Tabela 5	Valores da distribuição teórica $F_V(v)$ para cada amostra v	57
Tabela 6	Erro percentual ξ_{b_i} para o erro médio ϵ_{b_i} da estimativa do tempo de chegada em relação ao tempo efetivo Γ_{b_i}	63

LISTA DE ABREVIATURAS E SIGLAS

ITS	<i>Intelligent Transportation Systems</i>	2
APTS	<i>Advanced Public Transportation Systems</i>	2
GPS	<i>Global Positioning System</i>	2
AVL	<i>Automatic Vehicle Location</i>	2
ANN	<i>Artificial Neural Networks</i>	7
MAPE	<i>Mean Absolute Percentage Error</i>	8
SSNN	<i>State Space Neural Networks</i>	8
EKF	<i>Extended Kalman Filter</i>	8
APC	<i>Automatic Passenger Counters</i>	9
DoD	<i>United States Department of Defense</i>	12
GIS	<i>Geographic Information System</i>	13
SA	<i>Selective Availability</i>	14
MRU	Movimento Retilíneo Uniforme	36
API	<i>Application Programming Interface</i>	41
FDA	Função de Distribuição Acumulada	85
FMP	Função Massa de Probabilidade	85
FDP	Função Densidade de Probabilidade	86

LISTA DE SÍMBOLOS

D	Estatística de Kolmogorov-Smirnov.....	17
X	Vetor de variáveis de estado.....	23
u	Vetor de sinais de controle.....	23
W	Ruído no processo.....	23
A	Matriz de transição de estados.....	23
B	Matriz de controle.....	23
z	Vetor de medição.....	23
H	Matriz de medição.....	23
e	Ruído na medição.....	23
Q	Matriz de covariância do ruído do processo.....	23
R	Matriz de covariância do ruído de medição.....	23
\hat{X}^-	Estado estimado <i>a priori</i>	24
\hat{X}	Estado estimado <i>a posteriori</i>	24
y^-	Erro estimado <i>a priori</i>	24
y	Erro estimado <i>a posteriori</i>	24
P^-	Covariância do erro estimado <i>a priori</i>	24
P	Covariância do erro estimado <i>a posteriori</i>	24
K	Ganho de Kalman.....	24
x	Posição do ônibus.....	36
v	Velocidade do ônibus.....	36
w_x	Ruído de posição.....	36
w_v	Ruído de velocidade.....	36
ψ	Velocidade hipotética.....	37
X_p	Distância total de um percurso de ônibus urbano.....	38
T_h	Tempo médio histórico de um percurso de ônibus urbano ..	38
Ω_{BS}	Vetor de tempos hipotéticos de chegada em pontos de parada	39
Υ_b	Tempo de chegada estimado do ônibus em pontos de parada	39
V_c	Velocidade de cruzamento.....	42
BS	Vetor de posições de pontos de ônibus.....	42
b	Posição de um ponto de ônibus.....	42
S	Matriz de medidas de posição de ônibus.....	42
V	Matriz de medidas de velocidade de ônibus.....	42

E_{\max}	Erro máximo absoluto da estimativa de posição do ônibus .	48
Γ_b	Tempo efetivo de chegada do ônibus em pontos de parada .	49
ϵ_b	Erro médio da estimativa de tempo de chegada	49
ξ_b	Erro percentual da estimativa de tempo de chegada	50
E	Vetor de erros de estimação dos tempos de chegada do ônibus entre dois pontos de parada	54
ζ	Erro de estimação dos tempos de chegada do ônibus entre dois pontos de parada	54

SUMÁRIO

1 INTRODUÇÃO	1
1.1 PROPOSTA DO TRABALHO	3
1.2 OBJETIVOS	4
1.2.1 Objetivos Específicos	4
1.3 ORGANIZAÇÃO DO TRABALHO	4
2 TRABALHOS RELACIONADOS	7
3 FUNDAMENTAÇÃO TEÓRICA	11
3.1 LOCALIZAÇÃO AUTOMÁTICA DE VEÍCULOS	11
3.1.1 Tecnologia de Localização por GPS	12
3.1.2 Tecnologia de Transmissão de Dados	15
3.2 TESTE DE KOLMOGOROV-SMIRNOV	16
3.3 FILTRO DE KALMAN	21
3.3.1 Processo a Ser Estimado	23
3.3.2 Origem Computacional do Filtro de Kalman	23
3.3.3 Origem Probabilística do Filtro de Kalman	25
3.3.4 Algoritmo do Filtro de Kalman	25
3.3.5 Parametrização do Filtro de Kalman	28
3.4 SUMÁRIO	28
4 ANÁLISE ESTATÍSTICA DE DADOS DE POSIÇÃO E VELOCIDADE DE ÔNIBUS URBANO	29
4.1 ANÁLISE LONGITUDINAL DE DADOS DE MEDIÇÃO ..	29
4.2 ANÁLISE TRANSVERSAL DE DADOS DE MEDIÇÃO ...	32
4.3 SUMÁRIO	34
5 MODELO DE PREDIÇÃO DE TEMPOS DE CHE- GADA DE ÔNIBUS URBANO	35
5.1 MODELO PARA ESTIMAÇÃO DE POSIÇÃO E VELOCI- DADE DO ÔNIBUS	35
5.2 ALGORITMO PARA PREDIÇÃO DE TEMPOS DE CHE- GADA DE ÔNIBUS URBANO	38
5.3 SUMÁRIO	40
6 RESULTADOS NUMÉRICOS E SIMULAÇÕES	41
6.1 ESTIMAÇÃO DOS TEMPOS DE CHEGADA DO ÔNIBUS PARA O CENÁRIO I	41
6.1.1 Aplicação do Teste de Kolmogorov-Smirnov	42
6.1.2 Configuração do Filtro de Kalman	45
6.1.3 Tempos de Chegada Estimados	46

6.2 ESTIMAÇÃO DOS TEMPOS DE CHEGADA DO ÔNIBUS PARA O CENÁRIO II	55
6.2.1 Aplicação do Teste de Kolmogorov-Smirnov	56
6.2.2 Configuração do Filtro de Kalman	58
6.2.3 Tempos de Chegada Estimados	59
6.3 SUMÁRIO	67
7 CONCLUSÃO	69
7.1 TRABALHOS FUTUROS	70
REFERÊNCIAS	71
APÊNDICE A – Tabela para Funções de Distribuição Normal Padronizadas	79
APÊNDICE B – Probabilidade e Variáveis Aleatórias ..	83

1 INTRODUÇÃO

Qualquer pessoa que deseje viajar entre dois pontos terá que tomar algumas decisões como o tipo de transporte, a rota, o tempo de viagem, etc. Na maioria dos casos, o modo de viagem privado é o preferido pelo público por causa do conforto, conveniência e flexibilidade oferecidos pelo mesmo (CORPUZ, 2007). Entretanto, modos privados de transporte são mais caros e levam a congestionamentos maiores que reduzem a eficiência da infraestrutura de transporte e aumentam os tempos de viagem, a poluição do ar, e o consumo de combustível (PADMANABAN et al., 2010).

Uma maneira de resolver este problema seria fornecendo mais infraestrutura para acomodar o número crescente de veículos. Todavia, há um limite para esta solução e por isso outras alternativas precisam ser exploradas para atender à demanda crescente de tráfego (ZHU et al., 2011). Uma opção é um maior desenvolvimento dos sistemas de transporte público.

Transporte público é um serviço de transporte de pessoas compartilhado, disponível para utilização de toda a população. Modos de transporte coletivo incluem ônibus, bondes, trens, metrô, entre outros. Carros removidos das vias através da escolha pelo transporte público traduz-se em congestionamentos e tempos de viagem reduzidos, diminuindo consideravelmente o custo total com transporte. A opção pelo transporte público pode resultar ainda em uma redução significativa na emissão de carbono. Uma única pessoa que escolha andar 37 km por dia utilizando, por exemplo, um ônibus, gera uma redução na emissão de CO₂ de aproximadamente 2177 kg/ano (LI et al., 2011).

Muitas agências de trânsito ao redor do mundo têm desenvolvido cada vez mais seus sistemas de transporte público. Um problema chave no processo de desenvolvimento destes sistemas é o provimento de um serviço de transporte público adequado e eficiente (ABDELFATTAH; KHAN, 1998).

Para a operação eficiente de linhas de ônibus urbano, é importante conhecer a posição dos veículos em tempo real. Assim, por exemplo, torna-se possível realizar o controle dos instantes de partida em pontos de parada (DELGADO et al., 2009; KOEHLER; KRAUS; CAMPO-NOGARA, 2008) e implantar sistemas de informação sobre chegadas futuras para os usuários, melhorando a percepção da qualidade do serviço prestado (ABDEL-ATY, 2001; SCHWEIGER, 2003). Além disso, operadores de trânsito podem identificar veículos com atrasos significativos e

agir de maneira proativa (JEONG; RILETT, 2004).

A informação sobre os tempos de chegada de ônibus urbano é o componente vital de muitas aplicações de sistemas inteligentes de transporte (ITS, do inglês *Intelligent Transportation Systems*) (QUDDUS; OCHIENG; NOLAND, 2007). Neste contexto, sistemas de transporte público avançados (APTS, do inglês *Advanced Public Transportation Systems*) são uma das mais importantes aplicações de ITS (ZHU et al., 2011). Todavia, para estimar os tempos de chegada de ônibus, sistemas dinâmicos podem ser desenvolvidos utilizando dados coletados por tecnologias como o sistema de posicionamento global (GPS, do inglês *Global Positioning System*), que, nos últimos anos, se estabeleceu como uma importante tecnologia no fornecimento de dados de localização para aplicações de ITS (QUDDUS; OCHIENG; NOLAND, 2007).

Sistemas para localização automática de veículos (AVL, do inglês *Automatic Vehicle Location*), inseridos no contexto de APTS, têm sido adotados por muitas agências de trânsito, pois permitem o rastreamento eficaz de seus veículos em tempo real (JEONG; RILETT, 2004). Entretanto, enquanto a localização de um ônibus é relativamente simples, a previsão de quando este irá chegar em um determinado local é algo significativamente mais complexo (JEONG; RILETT, 2004).

Desde que a predição dos tempos de chegada do ônibus dependa de uma série de fatores (por exemplo, atrasos em interseções sinalizadas, incidentes, número de pontos de parada, etc.), condições estocásticas de tráfego ao longo do trajeto e a variação do número de passageiros nos pontos de parada, aumentam o nível de incerteza consideravelmente (ZHU et al., 2011).

Vários estudos têm sugerido diferentes técnicas para predição eficiente de tempos de chegada. Alguns deles incluem algoritmos de tempo real, algoritmos estatísticos para análise de dados históricos (análises de regressão e séries temporais, cálculo de médias e variâncias), abordagens baseadas em modelo (filtragem de Kalman) e aprendizado de máquina (redes neurais artificiais) (SEEMA; ALEX, 2009; SHALABY; FARHAN, 2002; JEONG; RILETT, 2004). Algoritmos estatísticos para análise de dados históricos, realizam a predição do tempo para um momento particular como o tempo de viagem médio para este mesmo período ao longo de diferentes dias. Abordagens de tempo real predizem o próximo intervalo de tempo de viagem como sendo o mesmo do tempo de viagem presente. Modelos de regressão são abordagens convencionais no que diz respeito à predição de tempos de viagem e predizem uma variável dependente baseando-se em uma função formada por um conjunto de variáveis independentes. Aprendizado de máquina

é uma das mais comuns técnicas de predição citadas para sistemas de tráfego por sua habilidade de resolver relações complexas não lineares. As técnicas baseadas em filtros de Kalman realizam estimações a partir de modelos fenomenológicos, que consideram propriedades estocásticas das perturbações do processo e ruídos de medição (ZHU et al., 2011).

1.1 PROPOSTA DO TRABALHO

Neste trabalho, propõe-se realizar a predição dos tempos de chegada do ônibus a pontos de parada por meio da abordagem baseada em modelo, utilizando o filtro de Kalman e análises de dados históricos. O trabalho é inspirado nas propostas de Dailey et al. (2000), Wall e Dailey (1999) e principalmente Cathey e Dailey (2003). O método desenvolvido é sistemático e claro de ser reproduzido, ao contrário dos trabalhos de Dailey, que omitem aspectos de implementação das técnicas descritas. São propostos algoritmos estatísticos para análise de dados históricos e um modelo dinâmico linear com ruídos, que considera somente os estados de posição e velocidade do ônibus. Além disso, os ruídos de processo são associados aos estados de posição e velocidade, não à aceleração, como acontece no trabalho de Cathey e Dailey (2003). Este modelo linear com ruídos, mesmo sendo mais simples, mostrou-se apropriado, confirmando a eficiência na tarefa de estimação dos tempos de chegada.

Atrasos em interseções sinalizadas, variação do número de passageiros para embarque/desembarque em pontos de parada, etc., são considerados como propriedades estocásticas das perturbações do processo. Propõe-se a utilização do filtro de Kalman para estimação dos estados de posição e velocidade do veículo ao longo do trajeto. Dados observados e dados históricos, utilizados para definição dos parâmetros do filtro Kalman, são obtidos através da simulação de dois cenários distintos implementados no *software* de simulações microscópicas Aimsun 6.1. São propostos dois métodos de análise estatística para extração de informações dos dados de medição de posição e velocidade: o método de análise longitudinal, e o método de análise transversal. É proposto também, um algoritmo para predição dos tempos de chegada do ônibus em pontos de parada. É proposta ainda, a aplicação do teste de ajustamento de Kolmogorov-Smirnov para identificar se os dados utilizados seguem uma distribuição normal, necessária para a aplicação do algoritmo do filtro de Kalman. Considera-se os dados observados como sendo oriundos de um GPS emulado no *software* Aimsun, associado a

um sistema AVL agregado ao ônibus.

1.2 OBJETIVOS

O objetivo geral deste trabalho é realizar um sistema que permita:

- a previsão de próximas chegadas de ônibus a pontos de parada;
- a estimativa de tempo de percurso para operadores a fim de tomada de medidas corretivas;
- o barateamento das tecnologias a serem empregadas através da substituição de trocas frequentes de dados pela estimação a partir de um número reduzido de amostras.

1.2.1 Objetivos Específicos

Os objetivos específicos podem ser divididos em:

- definição de métodos para o tratamento de dados de medição de posição e velocidade do ônibus;
- aplicar um teste de ajustamento para testar a distribuição estatísticas dos dados de medição de velocidade;
- aplicar filtros de Kalman para filtragem de erros inerentes à tecnologia de rastreamento e estimação dos estados de posição e velocidade do ônibus;
- desenvolver um algoritmo para predição dos tempos de chegada do ônibus a pontos de parada.

1.3 ORGANIZAÇÃO DO TRABALHO

No capítulo 2, são apresentados os trabalhos relacionados com o objetivo de expor as diferentes abordagens disponíveis para a solução do problema em questão, porém, com um foco maior na abordagem escolhida para o desenvolvimento deste trabalho.

No capítulo 3, o sistema AVL e seu principal componente, o GPS, são abordados. Os conceitos estatísticos e o teste de ajustamento

escolhido são definidos. Da mesma forma, são apresentados os conceitos relacionados à filtragem de Kalman, bem como seu algoritmo e suas origens, tanto computacional quanto probabilística.

No capítulo 4, são apresentados os dois métodos de análise estatística de dados históricos, desenvolvidos com o objetivo de extrair informações relevantes à configuração dos parâmetros do filtro de Kalman e definição da distribuição estatística dos dados observados ou históricos.

No capítulo 5, a modelagem matemática desenvolvida para representação do sistema é apresentada. Além disso, é mostrado o algoritmo idealizado para a estimação do tempo de chegada do ônibus a um ponto de parada qualquer.

No capítulo 6, apresenta-se os cenários de simulação desenvolvidos para geração de dados de medição. Além disso, são mostrados os resultados numéricos para os testes de ajustamento realizados aos dados observados, bem como os passos necessários para a definição dos parâmetros do filtro de Kalman. Da mesma forma, são apresentados os resultados relacionados à aplicação do modelo proposto para a estimação do tempo de chegada do ônibus em cinco pontos de parada dos dois cenários desenvolvidos.

Finalmente, no capítulo 7 são apresentadas algumas considerações finais relacionadas à realização deste trabalho, bem como propostas para trabalhos futuros.

2 TRABALHOS RELACIONADOS

Sistemas AVL são frequentemente utilizados em sistemas de transporte público, principalmente os que envolvem ônibus urbano. O elemento sensor destes sistemas é baseado em um GPS responsável por fornecer dados em tempo real da posição do veículo. Aplicando um método estatístico específico para tratamento destes dados, é possível desenvolver modelos eficientes para estimação de tempos de chegada/partida de ônibus urbano. Segundo Li et al. (2011), os modelos mais comuns utilizados para predição de tempos de chegada de ônibus podem ser classificados em: métodos matemáticos de predição (que podem envolver algoritmos de tempo real, modelos de regressão, algoritmos estatísticos e etc.), redes neurais artificiais (ANN, do inglês *Artificial Neural Networks*), e filtros de Kalman com análise de dados históricos.

Lin e Zeng (1999) desenvolveram um algoritmo matemático para prover informações em tempo real sobre tempos de chegada de ônibus urbano. São consideradas informações sobre tempos programados de chegada, dados de localização do veículo, a diferença entre o tempo atual de chegada e o tempo programado de chegada, e o tempo de espera em paradas de verificação de tempo, presentes no algoritmo. Os dados de posição do ônibus são obtidos através de um sistema AVL. O algoritmo foi principalmente desenvolvido para sistemas de informação em áreas rurais onde não há congestionamentos, não considerando assim, o tráfego e os tempos de permanência em cada parada.

Em Li et al. (2011), é apresentada uma abordagem estatística para predição dos tempos de chegada de ônibus urbano baseada em um sistema de gerenciamento de informações de tráfego. São considerados alguns fatores que afetam o tempo de viagem do veículo, como o tempo de partida, a localização atual do ônibus, o número de interseções, a demanda de passageiros a cada ponto de parada, o estado do tráfego da rede urbana, etc. Os tempos de chegada do veículo são descritos através de um modelo linear e os parâmetros do modelo são obtidos através dos dados históricos de tempos de chegada. Os dados de posição de ônibus são obtidos por GPS.

Um método estatístico para predição dos tempos de chegada do ônibus que incorpora explicitamente os atrasos em pontos de parada e interseções sinalizadas associadas aos tempos totais de viagem do ônibus é apresentado por Zhu et al. (2011). O cálculo do erro percentual absoluto médio (MAPE, do inglês *Mean Absolute Percentage Error*) é

utilizado como uma medida da precisão da estimativa. Um GPS é utilizado para obtenção dos dados de posição do ônibus urbano.

Chien, Ding e Wei (2002) desenvolveram um modelo de rede neural artificial para predição dos tempos de chegada de ônibus urbano utilizando o algoritmo *backpropagation*, que é o algoritmo mais empregado para problemas de transporte. Todavia, como este algoritmo é inadequado para aplicações *online*, os autores desenvolveram um fator de ajuste com o objetivo de modificar a predição do tempo de chegada através da utilização de dados observados em tempo real. Entretanto, tempos de permanência em pontos de parada e tempos de chegada programados não são considerados. Um gerador de dados é utilizado na predição do tempo de chegada do ônibus. A avaliação de desempenho é realizada através de experimentos efetuados no simulador microscópico de tráfego CORSIM.

Jeong e Rilett (2004) avaliam o desempenho de modelos baseados em dados históricos, de regressão e ANN, para predição dos tempos de chegada do ônibus, considerando congestionamentos, tempos de chegada programados e tempos de permanência em pontos de parada. Mostram que o modelo ANN apresenta desempenho superior aos outros dois modelos. Os dados de posição do ônibus são obtidos através de um sistema AVL.

Liu et al. (2006) desenvolveram um modelo híbrido baseado em redes neurais de espaço de estados (SSNN, do inglês *State Space Neural Networks*) e filtro de Kalman estendido (EKF, do inglês *Extended Kalman Filter*). O modelo SSNN precisa de um conjunto extenso de dados para treinamento *offline*. O modelo de filtro de Kalman estendido foi desenvolvido para treinamento do SSNN. Os dados são provenientes de um sistema real.

Os primeiros a trabalhar com filtro de Kalman para predição de tempos de chegada de ônibus urbano foram Wall e Dailey (1999). O algoritmo de predição utiliza uma combinação de dados históricos e dados oriundos do sistema AVL para produzir uma distribuição dos tempos de viagem. É dividido em dois componentes: o rastreador (*tracking*), que emprega o filtro de Kalman; e o preditor (*prediction*), que aplica estimações estatísticas. Tempos de permanência em pontos de parada, interseções sinalizadas, congestionamentos e afins, são considerados como ruídos na trajetória do veículo.

Em Shalaby e Farhan (2002), é proposto outro preditor de tempos de viagem utilizando filtros de Kalman. Em seu modelo, o trajeto é dividido em *links* e são consideradas informações sobre o número de passageiros a cada ponto de parada de cada *link*. Dois algoritmos de

filtragem de Kalman são desenvolvidos para predição dos tempos de percurso nos *links* e dos tempos de permanência em pontos de ônibus, separadamente. Neste último, os tempos de permanência são estimados somente nos *check-points* presentes no algoritmo, não a cada parada. Os dados de posição do veículo são obtidos através de um sistema AVL, enquanto que as informações sobre passageiros são adquiridas através do sistema de contagem automática de passageiros (APC, do inglês *Automatic Passenger Counters*).

Cathey e Dailey (2003) desenvolveram um modelo para predição de tempos de chegada de ônibus baseando-se em filtros de Kalman, dividido em três componentes distintos: o rastreador (*tracker*), baseado no sistema AVL; o filtro (*filter*), que utiliza a técnica de filtragem de Kalman para transformar uma sequência de medidas provenientes do sistema AVL em estimativas do estado dinâmico do veículo; e o preditor (*predictor*), que realiza predições precisas sobre o comportamento do veículo em trânsito através de estimações estatísticas. O movimento Browniano é utilizado na representação da aleatoriedade na aceleração do veículo. Congestionamentos, interseções sinalizadas, número de pontos de parada, tempos de permanência nestes pontos, entre outros, são considerados como sendo ruídos do processo.

O modelo dinâmico linear com ruídos, desenvolvido para representação da movimentação do ônibus no trabalho de Cathey e Dailey (2003), foi a principal motivação para a utilização da abordagem de modelos baseados em filtros de Kalman e análise estatística de dados históricos. Além disso, segundo Shalaby e Farhan (2002), modelos baseados em filtros de Kalman, superam modelos baseados em métodos matemáticos de predição ou ANN no contexto de precisão da predição dos tempos de chegada. Então, como o objetivo é de obter boa precisão em predições de tempos chegada do ônibus, esta abordagem foi escolhida.

Na Tabela 1 pode ser vista uma comparação das técnicas aqui apresentadas com referências e comentários.

Tabela 1 – Comparação de técnicas para predição de tempos de chegada de ônibus urbano.

Autor(es)	Origem dos Dados	Modelo	Comentários
Lin e Zeng (1999)	Sistema AVL	Métodos matemáticos de predição	Considera informações sobre tempos programados de chegada, dados de localização do veículo, a diferença entre os tempos atual e programado de chegada, etc.
Li et al. (2011)	GPS	Métodos matemáticos de predição	Baseia-se em um sistema de gerenciamento de informações de tráfego. São considerados fatores externos que podem afetar os tempos de viagem do veículo.
Zhu et al. (2011)	GPS	Métodos matemáticos de predição	Incorpora atrasos em pontos de parada e interseções sinalizadas associadas aos tempos totais de viagem do ônibus.
Chien, Ding e Wei (2002)	Dados simulados no CORSIM	ANN	Utiliza o algoritmo <i>backpropagation</i> com um fator de ajuste. Tempos de permanência em pontos de parada e tempos de chegada programados não são considerados.
Jeong e Rilett (2004)	Sistema AVL	ANN	Compara o modelo ANN com o modelo de regressão e o modelo baseado em dados históricos. O modelo ANN apresenta melhores resultados.
Liu et al. (2006)	GPS	SSNN e EKF	Modelo híbrido baseado em SSNN e EKF.
Wall e Dailey (1999)	Sistema AVL	Filtros de Kalman e análise estatística de dados históricos	Sistema dividido em dois componentes: rastreador, que emprega o filtro de Kalman; e preditor, que aplica estimações estatísticas.
Shalaby e Farhan (2002)	Sistema AVL e APC	Filtros de Kalman e análise estatística de dados históricos	Utiliza dois modelos de filtro de Kalman para estimar tempos de chegada baseando-se em dados oriundos dos sistemas AVL e APC.
Cathey e Dailey (2003)	Sistema AVL	Filtros de Kalman e análise estatística de dados históricos	Sistema dividido em três componentes: rastreador (sistema AVL), filtro (filtro de Kalman), e preditor (estimções estatísticas).

3 FUNDAMENTAÇÃO TEÓRICA

De acordo com a proposta apresentada no Capítulo 1 (Seção 1.1), busca-se desenvolver um sistema que utilize um filtro de Kalman para estimação do estado dinâmico do veículo a partir de um modelo fenomenológico, dados históricos e medições advindas de GPS. Portanto, descreve-se neste capítulo os conceitos básicos aplicados no processo de pesquisa e desenvolvimento deste trabalho. Inicialmente, é definido o sistema para Localização Automática de Veículos (AVL) e sua principal tecnologia de localização, o sistema de posicionamento global (GPS). As principais fontes de incerteza do GPS são discutidas. Em seguida, são abordados os fundamentos de uma ferramenta para a realização de testes de ajustamento, o teste de Kolmogorov-Smirnov, aplicado com o objetivo de avaliar a hipótese de que os dados históricos de velocidade do ônibus seguem uma distribuição normal. Finalmente, é apresentada a teoria do filtro de Kalman, seu algoritmo e suas origens, tanto computacional quanto probabilística. O filtro Kalman é utilizado para estimação do estado dinâmico do ônibus, necessário para estimação dos tempos de chegada do mesmo a pontos de parada ao longo do percurso. No Apêndice B, é realizada uma revisão dos conceitos estatísticos, como probabilidade e variáveis aleatórias, relacionados com o desenvolvimento deste trabalho.

3.1 LOCALIZAÇÃO AUTOMÁTICA DE VEÍCULOS

Sistemas de localização automática de veículos (AVL) são sistemas informáticos computacionais empenhados na tarefa de rastrear veículos (CASEY et al., 1998). São utilizados no trânsito, frotas de caminhões, veículos oficiais e para propósitos militares. Todavia, é crescente a utilização desses sistemas no rastreamento de veículos de trânsito dado os vários benefícios que esta tecnologia proporciona (CASEY et al., 1998). Dentre vários benefícios concedidos pelo uso desta tecnologia, pode-se destacar:

- O sistema AVL coleta informações em tempo real que podem estar disponíveis ao público;
- Melhora a confiabilidade do cronograma de viagem;
- Reduz custos de operação e manutenção;

- Melhora a eficiência do serviço;
- Reforça a segurança;
- As agências de trânsito podem responder mais rapidamente a situações de emergência.

A tecnologia AVL foi utilizada para rastrear veículos em trânsito pela primeira vez em Londres na Inglaterra, no final da década de 1950 (TURNBULL, 1993). Esse sistema é baseado em duas tecnologias: localização e transmissão de dados. A tecnologia de localização é utilizada para medir a posição em tempo real de um veículo qualquer. A tecnologia de transmissão de dados é utilizada para transmitir a informação a uma central de localização (CASEY et al., 1998).

3.1.1 Tecnologia de Localização por GPS

A principal tecnologia utilizada na localização de veículos em trânsito é o sistema de posicionamento global (GPS - *Global Positioning System*). A tecnologia GPS determina a posição do veículo utilizando os sinais transmitidos por mais de 28 satélites. O GPS funciona em todos os lugares alcançáveis pelos satélites, e isso torna o GPS uma tecnologia mais robusta que as outras tecnologias de localização disponíveis (JEONG, 2004). Entretanto, sinais de satélite não alcançam o subsolo e podem ter seu sinal interrompido por construções e afins. Onde esses problemas acontecem, outras tecnologias de localização podem ser utilizadas de forma a complementar o sinal do GPS (JEONG, 2004; TAYLOR et al., 2006).

O GPS é uma valiosa fonte de dados para cientistas, pois permite que pesquisadores em todo o mundo conheçam a localização de construções, veículos, pessoas e outros objetos de interesse a um custo relativamente baixo. Seu desenvolvimento tem suas raízes no Departamento de Defesa dos Estados Unidos da América (DoD) no início dos anos 60. O DoD desenvolveu um satélite baseado em sistemas de radioposicionamento, de forma a obter apoio com boa precisão para navegação e posicionamento de submarinos de mísseis balísticos (DEFENSE, 1996). O sistema operacional total do GPS possui mais de 28 satélites orbitando a uma distância de mais de 20000 km acima da Terra. A constelação de órbitas ilustrada na Figura 1 foi projetada para que no mínimo quatro satélites estejam visíveis em qualquer lugar da Terra em qualquer momento. Cada satélite transmite sinais de rádio que podem ser utilizados para calcular uma posição (JOHN et al., 2003).



Figura 1 – A constelação de satélites para GPS. FONTE: (JOHN et al., 2003)

Segundo John et al. (2003), um receptor GPS, basicamente, coleta um par de coordenadas (latitude e longitude) para a localização na qual o usuário se encontra naquele momento em particular. Este local é percebido como um ponto quando colocado em um mapa. É importante salientar que o GPS é mais frequentemente utilizado para coletar coordenadas de uma localização individual, o qual é percebido como pontos em um sistema de informação geográfica (GIS - *Geographic Information System*). No entanto, alguns receptores GPS também podem ser utilizados para coletar e gerar dados em linha e áreas. Isto pode ser feito de muitas maneiras, mas uma delas seria a de coletar pontos discretos ao longo do comprimento de uma linha ou a cada ângulo de uma área (polígono) específica. Os pontos podem então ser inseridos em um GIS e conectados manualmente por segmentos de reta para formar a linha ou polígono.

Por causa das vantagens que o GPS proporciona, muitos estudos sobre a coleta de dados utilizando esta tecnologia têm sido realizados. O GPS é utilizado para coletar dados de tempo de viagem, velocidade, escolha de rotas, posição, tempo, entre outros. Esses estudos têm mostrado que a utilização do GPS para coleta de dados é mais fácil e precisa do que nos métodos tradicionais (JEONG, 2004). Entretanto, no GPS, também existem fontes de erros.

Os erros que afetam as observações GPS possuem, geralmente, três origens: erros dependentes dos satélites, erros dependentes da

antena-receptor e erros dependentes do meio de propagação (KAPLAN; HEGARTY, 2006; GUERREIRO, 2005; MONICO, 2000).

Os **erros dependentes dos satélites** podem ser divididos em três, sendo:

1. **Erros nos relógios dos satélites.** Embora os relógios dos satélites sejam consideravelmente precisos (cada satélite contém quatro relógios atômicos, dois de rubídio e dois de césio), estes não são perfeitos. Apenas um nanosegundo de erro, ou seja 1 ns, resulta em um erro de cerca de 30 cm na medição da distância para um satélite.
2. **Erros nas efemérides.** A precisão da posição depende da exatidão do conhecimento da localização dos satélites. O DoD, coloca cada satélite em uma órbita precisa, fazendo com que esta seja previsível por um modelo matemático rigoroso. No entanto, o pouco conhecimento do campo gravitacional terrestre, as forças gravitacionais da Lua e do Sol, e o atrito remanescente da atmosfera terrestre, bem como a pressão das radiações solares nos satélites, provocam variações em suas órbitas, por isso a necessidade da constante monitoração pelas estações de rastreamento de satélites na Terra.
3. **Acesso seletivo (SA).** Consiste na manipulação da mensagem de navegação de modo a degradar a informação inerente ao relógio do satélite e às efemérides radiodifundidas. O SA foi, entretanto, removido em 1 de Maio de 2000. O código geral foi idealizado para dar uma precisão de 30 metros, porém sua precisão original era reduzida para 100 metros.

Os **erros dependentes da antena-receptor** ocorrem nos relógios dos receptores, de forma semelhante aos erros provocados pelos relógios dos satélites. Também são divididos em três:

1. **Multi-Trajeto.** Assume-se que o sinal do satélite viaja diretamente desde o satélite até a antena do receptor. Entretanto, existem sinais refletidos, provocados por objetos que se encontram próximos à antena, interferindo no sinal verdadeiro. Este efeito é chamado de multi-trajeto e afeta apenas medições de alta precisão. Sua magnitude é de aproximadamente 50 cm.
2. **Erros causados pela variação do centro de fase da antena.** Estes erros ocorrem em função da construção da antena.

Dependem do ângulo e direção do sinal observado. Estas variações podem ser de até alguns centímetros.

3. **Ruídos do receptor.** O receptor GPS tem suas limitações. Está limitado à sua própria precisão, ou seja, ao desvio padrão associado a cada medição.

Os **erros dependentes do meio de propagação** podem ser divididos em dois:

1. **Atrasos ionosféricos.** Para calcular a distância de um satélite, mede-se o tempo que o sinal leva para chegar ao receptor e multiplica-se esse tempo pela velocidade da luz. O problema é que a velocidade da luz varia sob as condições atmosféricas. A camada mais alta da atmosfera, a ionosfera, contém partículas “carregadas” que atrasam o código e adiantam a fase. A magnitude deste efeito é maior durante o dia do que a noite. Os atrasos ionosféricos não modelados podem afetar a precisão em até 10 metros.
2. **Atrasos troposféricos.** Ao passar pela camada mais baixa da atmosfera, a troposfera, o sinal também sofre um atraso na fase e no código. Este atraso é causado por dois componentes: um componente seco e um úmido. O principal problema relaciona-se com o componente úmido (vapor de água).

3.1.2 Tecnologia de Transmissão de Dados

Jeong (2004) afirma que a informação sobre posição, independente da tecnologia de localização adotada, é geralmente armazenada no veículo em trânsito por um período específico de tempo. Esta informação é retransmitida ao centro de envio em sua forma bruta ou processada no próprio veículo. As duas tecnologias de transmissão de dados mais comuns são **votação e relatórios de exceção**.

Com a tecnologia de **votação**, o computador no centro de envio pergunta a cada ônibus sua localização em intervalos regulares. A precisão da localização é uma função de quantas vezes o veículo em trânsito é monitorado. Entretanto, o número de frequências de rádio disponíveis em áreas urbanas é limitado. Por esta razão, muitas agências de trânsito escolhem a tecnologia de relatórios de exceção (CASEY et al., 1998).

Pelo método de **relatórios de exceção**, cada ônibus relata sua localização em alguns locais específicos. Nesta tecnologia, faz-se necessário conhecer não somente a posição do veículo, mas também sua posição programada. Muitas agências de trânsito combinam os métodos de votação e relatórios de exceção buscando melhores resultados (JEONG, 2004).

A tecnologia AVL apresentada fornece dados brutos sobre a localização do veículo. Para embasar o desenvolvimento de técnicas estatísticas para o tratamento e filtragem destes dados, serão discutidas a seguir as técnicas utilizadas neste trabalho para a construção do estimador de posição e tempos de chegada dos ônibus a pontos de parada.

3.2 TESTE DE KOLMOGOROV-SMIRNOV

Frequentemente, faz-se necessário o teste de hipóteses sobre a distribuição de uma população. Se o objetivo deste teste é demonstrar a semelhança entre a distribuição de um conjunto de amostras qualquer e uma distribuição teórica específica, o teste é chamado de **teste de ajustamento** (em inglês, *test of goodness of fit*) (MASSEY, 1951).

Em alguns destes testes, deseja-se analisar a hipótese nula H_0 de que dois vetores de amostras aleatórias y_1 e y_2 procedem de uma mesma distribuição contínua. A hipótese alternativa é que estas amostras são de distribuições contínuas diferentes. Estes testes são chamados de testes não-paramétricos ou de distribuição livre. Um dos mais amplamente utilizados é o teste de **Kolmogorov-Smirnov** (STEPHENS, 1970).

O teste de Kolmogorov-Smirnov calcula a probabilidade de que duas distribuições sejam a mesma. Utiliza a métrica da diferença máxima absoluta entre duas funções de distribuição acumulada. É aplicável apenas para populações com distribuição acumulada contínua (DAILEY et al., 2000).

No Apêndice B, é realizada uma revisão dos conceitos estatísticos, como probabilidade e variáveis aleatórias, relacionados com o teste de Kolmogorov-Smirnov.

Supondo um conjunto de amostras com valores x_1, x_2, \dots, x_n observados a partir de uma população X , é possível construir um função de distribuição acumulada observada, organizando os valores das amostras em ordem crescente, denotados aqui por $x_{(1)}, x_{(2)}, \dots$; e determinando a função de distribuição de X em $x_{(1)}, x_{(2)}, \dots$, denotada por $S_n[x_{(1)}], S_n[x_{(2)}], \dots$, oriunda da relação $S_n(x) = i/n$, onde i é o número de observações menores ou iguais a x e n é o número total de amostras

(SOONG, 2004).

Segundo Massey (1951), Lampariello (2000) e Press et al. (2002), para comparar um conjunto de dados $S_n(x)$ com uma função de distribuição acumulada contínua $F_X(x)$ conhecida, o teste estatístico de Kolmogorov-Smirnov será:

$$D = \max_{-\infty < x < \infty} |S_n(x) - F_X(x)|. \quad (3.1)$$

Para comparar duas diferentes funções de distribuição acumulada $S_{n_1}(x)$ e $S_{n_2}(x)$, a estatística de Kolmogorov-Smirnov será:

$$D = \max_{-\infty < x < \infty} |S_{n_1}(x) - S_{n_2}(x)|. \quad (3.2)$$

Na Figura 2 são apresentados os gráficos para a estatística D do teste de Kolmogorov-Smirnov. Valores de uma distribuição medidas em x são comparados com uma distribuição teórica no qual sua distribuição de probabilidades acumulada é plotada como $F_X(x)$. A função de distribuição acumulada $S_n(x)$ é construída. D representa a maior distância entre as duas distribuições acumuladas.

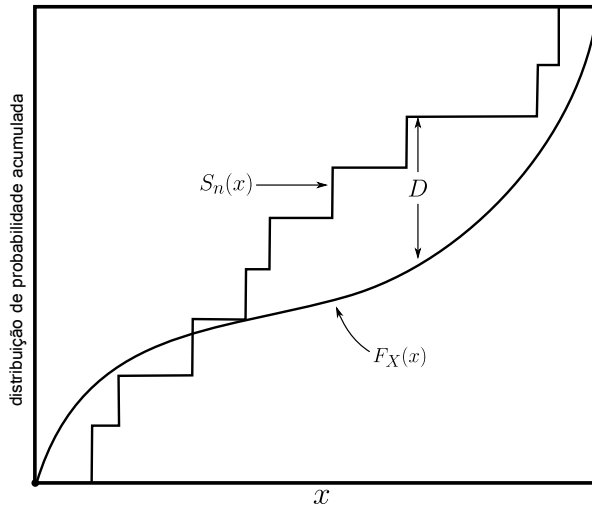


Figura 2 – Estatística D do teste de Kolmogorov-Smirnov. Baseado em Press et al. (2002).

Na Tabela 2 podem ser vistos valores críticos $d_\alpha(n)$ da distribui-

ção de D para algumas quantidades de amostras. Por exemplo, num nível de significância 0,20, o valor crítico de D para $n = 10$ é 0,322; isto significa que em 20% das variáveis aleatórias a diferença máxima absoluta entre duas funções de distribuição acumulada será de no mínimo 0,322. Aqui, os valores de $d_\alpha(n)$ são definidos por $P(D > d_\alpha(n)) = \alpha$ (MASSEY, 1951).

Tabela 2 – Valores críticos $d_\alpha(n)$ para diferença absoluta máxima entre duas funções de distribuição acumulada.

Amostras (n)	Nível de Significância (α)				
	0,20	0,15	0,10	0,05	0,01
5	0,446	0,474	0,510	0,565	0,669
10	0,322	0,342	0,368	0,410	0,490
15	0,266	0,283	0,304	0,338	0,404
20	0,231	0,246	0,264	0,294	0,356
25	0,21	0,22	0,24	0,27	0,32
30	0,19	0,20	0,22	0,24	0,29
35	0,18	0,19	0,21	0,23	0,27
> 35	$\frac{1,07}{\sqrt{n}}$	$\frac{1,14}{\sqrt{n}}$	$\frac{1,22}{\sqrt{n}}$	$\frac{1,36}{\sqrt{n}}$	$\frac{1,63}{\sqrt{n}}$

Na Figura 3 pode ser visto o método gráfico para a aplicação do teste de Kolmogorov-Smirnov. A curva contínua representa a distribuição acumulada $F_X(x)$. As curvas tracejadas estão a uma distância $\pm d_\alpha(n)$ da distribuição $F_X(x)$. A linha quebrada representa a função de distribuição acumulada de um determinado grupo de dados observados $S_n(x)$. Como $S_n(x)$ excedeu $d_\alpha(n)$, a hipótese H_0 de que as duas funções procedem de uma mesma distribuição foi rejeitada (MASSEY, 1951).

Em Soong (2004) é apresentado um “passo a passo” para a execução do teste de Kolmogorov-Smirnov para uma função de distribuição $S_n(x)$ de um conjunto de dados qualquer e uma função de distribuição teórica $F_X(x)$. Resume-se em:

1. Passo 1: organizar os valores x_1, x_2, \dots, x_n das amostras em ordem crescente e nomeá-los por $x_{(1)}, x_{(2)}, \dots, x_{(n)}$.
2. Passo 2: determinar a função de distribuição dos valores observados $S_n(x)$ para cada $x_{(i)}$ através da relação $S_n(x) = i/n$.
3. Passo 3: determinar a função de distribuição teórica $F_X(x)$ em cada $x_{(i)}$ utilizando a distribuição em hipótese. Os parâmetros da

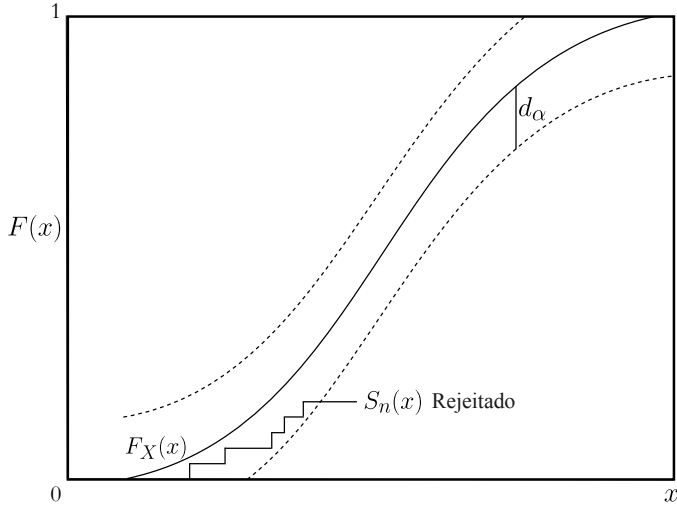


Figura 3 – Método gráfico para aplicação do teste de Kolmogorov-Smirnov. Baseado em Massey (1951).

distribuição, como média μ e variância σ^2 , podem ser estimados dos dados se necessário.

4. Passo 4: calcular $D = \max_{i=1}^n |S_n(x_{(i)}) - F_X(x_{(i)})|$.
5. Passo 5: escolher um valor para α e determinar, através da Tabela 2, o valor de $d_\alpha(n)$.
6. Passo 6: rejeitar a hipótese H_0 de que as duas funções de distribuição acumulada procedem de uma mesma distribuição se $D > d_\alpha(n)$. Caso contrário, hipótese H_0 é aceita.

Em Soong (2004) é apresentado ainda um exemplo que ilustra com clareza a aplicação do teste de Kolmogorov-Smirnov.

No exemplo, há 10 medidas de resistência à tração de um determinado material. Os dados são

$$X = [30,1; 30,5; 28,7; 31,6; 32,5; 29,0; 27,4; 29,1; 33,5; 31,0].$$

Baseando-se neste grupo de dados, deseja-se testar a hipótese nula H_0 de que a resistência à tração siga uma distribuição normal com nível de significância de até 5%.

O primeiro passo é reorganizar os dados em forma crescente. Assim, $x_{(1)} = 27,4$; $x_{(2)} = 28,7$;...; $x_{(10)} = 33,5$. Então, determina-se $S_n(x)$ através da relação $S_n(x) = i/n$. Portanto,

$$S_1(27,4) = \frac{1}{10} = 0,1; S_2(28,7) = \frac{2}{10} = 0,2; \dots; S_{10}(33,5) = \frac{10}{10} = 1.$$

No que diz respeito à função de distribuição teórica $F_X(x)$, a média e a variância são estimadas da seguinte maneira:

$$\hat{\mu} = \frac{1}{10} \sum_{i=1}^{10} x_i = 30,3,$$

$$\hat{\sigma}^2 = \frac{1}{10} \sum_{i=1}^{10} (x_i - 30,3)^2 = 3,14.$$

Os valores de $F_X(x)$ são encontrados baseando-se na distribuição $N(\hat{\mu} = 30,3; \hat{\sigma}^2 = 3,14)$ aplicando:

$$F_X(x) = F_Z \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right). \quad (3.3)$$

Por exemplo, com o auxílio da Tabela A.1, para a variável aleatória normal padronizada Z , têm-se

$$\begin{aligned} F_X(27,4) &= F_Z \left(\frac{27,4 - 30,3}{\sqrt{3,14}} \right) = F_Z(-1,64) = 1 - F_Z(-1,64) \\ &= 1 - 0,9495 = 0,0505, \end{aligned}$$

$$F_X(28,7) = F_Z \left(\frac{28,7 - 30,3}{\sqrt{3,14}} \right) = F_Z(-0,90) = 1 - 0,8159 = 0,1841,$$

e assim por diante.

Para determinar D , é construtivo plotar $S_n(x)$ e $F_X(x)$ em função de x , como apresentado na Figura 4. Vê-se claramente a partir da figura que a diferença máxima entre $S_n(x)$ e $F_X(x)$ ocorre em $x = x_{(4)} = 29,1$. Portanto:

$$D = |S_N(29,1) - F_X(29,1)| = 0,4 - 0,2483 = 0,1517.$$

Sendo $\alpha = 0,05$ e $n = 10$, a Tabela 2 fornece

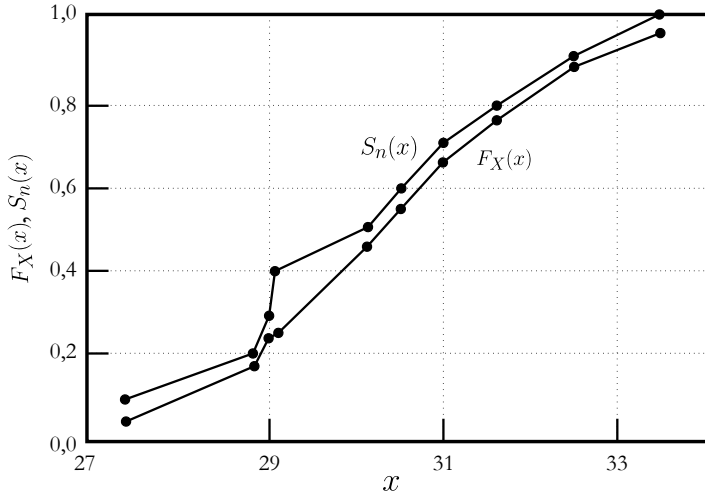


Figura 4 – $S_n(x)$ e $F_X(x)$ do exemplo. Baseado em Soong (2004).

$$d_{0,05}(10) = 0,41.$$

Desde que $D < d_{0,05}(10)$, a hipótese H_0 é verdadeira. Portanto, aceita-se a distribuição normal $N(30,3; 3,14)$ com até 5% de nível de significância.

3.3 FILTRO DE KALMAN

Das inúmeras ferramentas matemáticas que podem ser utilizadas para estimações estocásticas a partir de medições de sensores com ruídos, uma das mais conhecidas é o Filtro de Kalman. Em 1960, Rudolph E. Kalman publicou seu famoso artigo descrevendo uma solução recursiva para o problema de filtragem linear de dados discretos (KALMAN, 1960). Segundo Maybeck (1979), o filtro de Kalman é simplesmente um algoritmo recursivo ótimo para processamento de dados. Desde sua introdução, o filtro de Kalman tem sido o assunto de extensivas pesquisas e aplicações, principalmente na área de navegação autônoma e assistida.

Welch e Bishop (2001) descrevem que o filtro de Kalman é essencialmente um conjunto de equações matemáticas que implementam um

estimador do tipo preditor-corretor que é ótimo no sentido de que minimiza a covariância estimada do erro. Maybeck (1979) diz ainda que o filtro de Kalman pode incorporar todas as informações conhecidas do sistema, processando todas as medições disponíveis independente de sua precisão, para estimar o valor atual da variável de interesse através do uso do conhecimento do sistema, do conhecimento sobre a dinâmica do dispositivo de medição, da descrição estatística dos ruídos do sistema, dos erros de medição, da incerteza dos modelos dinâmicos e qualquer informação disponível sobre a condição inicial da variável de interesse.

Frequentemente, as variáveis de interesse não podem ser medidas diretamente, e algum meio de inferir esses valores, a partir de dados disponíveis, precisa ser definido. Esta inferência é complicada pelo fato que as relações entre as muitas variáveis de estado e saídas de medição são conhecidas apenas com algum grau de incerteza. Além disso, qualquer medição será corrompida por algum grau de ruído e imprecisões do dispositivo de medição. Assim, faz-se necessário prover um meio de extrair informações valiosas a partir de um sinal ruidoso. O filtro de Kalman combina todos os dados medidos disponíveis, mais o conhecimento *a priori* do sistema e dos dispositivos de medição, para produzir uma estimativa das variáveis desejadas de maneira que o erro de estimação seja minimizado estatisticamente. Em outras palavras, se executássemos alguns diferentes filtros muitas vezes para uma mesma aplicação, então, o resultado médio do filtro de Kalman seria melhor que a média dos resultados dos outros filtros (MAYBECK, 1979).

Conceitualmente, o que qualquer tipo de filtro tenta fazer é obter uma estimativa “ótima” a partir de dados fornecidos por um ambiente ruidoso. Existem muitas maneiras de realizar esse objetivo. Um deles seria o ponto de vista Bayesiano, onde o filtro propaga a densidade de probabilidade condicional, condicionado ao conhecimento atual dos dados oriundos do dispositivo de medição. O filtro de Kalman realiza esta propagação da densidade de probabilidade condicional para problemas nos quais o sistema pode ser descrito através de um modelo linear e no qual o sistema e os ruídos de medição são Brancos ou Gaussianos (MAYBECK, 1979).

Nesta seção, serão apresentados o algoritmo do filtro de Kalman e suas equações, bem como a origem computacional e probabilística deste filtro. Esta seção é baseada nos trabalhos de Welch e Bishop (2001), Kalman e Bucy (1961), Kalman (1960), Jazwinski (1970), Negenborn (2003), Ristic, Arulampalam e Gordon (2004), Meinhold e Singpurwalla (1983), Brown e Hwang (1992), Dailey et al. (2000), Cathey e Dailey

(2003) e Maybeck (1979).

3.3.1 Processo a Ser Estimado

O filtro de Kalman aborda o problema geral de tentar estimar o estado $X \in \mathbb{R}^n$ de um processo controlado em tempo discreto, governado por uma equação à diferença estocástica linear:

$$X_k = AX_{k-1} + Bu_{k-1} + W_{k-1}, \quad (3.4)$$

com uma medição $z \in \mathbb{R}^m$ dada por :

$$z_k = HX_k + e_k, \quad (3.5)$$

onde a matriz $n \times n$ A da Equação (3.4) é a matriz de transição de estados, B é a matriz de controle, X_k é o vetor que representa as variáveis de estado do sistema a cada instante de tempo k , e u é o vetor de sinais de controle. Na Equação (3.5), H é a matriz de medição, ou seja, relaciona o estado a medição z_k . As variáveis aleatórias W_k e e_k representam os ruídos de processo e medição, respectivamente. São definidas como variáveis independentes, brancas, com distribuição normal de probabilidade:

$$p(W) \sim N(0, Q), \quad (3.6)$$

$$p(e) \sim N(0, R), \quad (3.7)$$

onde Q é a matriz de covariância do ruído do processo, e R a matriz de covariância do ruído de medição. estas matrizes podem ser constantes ou variáveis a cada instante de tempo ou medição.

3.3.2 Origem Computacional do Filtro de Kalman

O estado estimado *a priori* no instante de tempo k , dado conhecimento do processo anterior a este instante, é definido por $\hat{X}_k^- \in \mathbb{R}^n$, enquanto que o estado estimado *a posteriori* no instante de tempo k , dada uma medição z_k , é definido por $\hat{X}_k \in \mathbb{R}^n$. Então, pode-se definir

os erros estimados *a priori* e *a posteriori*, respectivamente, como sendo:

$$y_k^- \equiv X_k - \hat{X}_k^-, \quad (3.8)$$

$$y_k \equiv X_k - \hat{X}_k. \quad (3.9)$$

A covariância do erro estimado *a priori* é dada por:

$$P_k^- = E[y_k^- y_k^{-T}]. \quad (3.10)$$

A covariância do erro estimado *a posteriori* será:

$$P_k = E[y_k y_k^T]. \quad (3.11)$$

Na derivação das equações do filtro de Kalman, o primeiro passo é encontrar uma equação que compute um estado estimado *a posteriori* \hat{X}_k como uma combinação linear de uma estimativa *a priori* \hat{X}_k^- e uma diferença ponderada entre uma medição real z_k e uma previsão de medição $H\hat{X}_k^-$, desta forma:

$$\hat{X}_k = \hat{X}_k^- + K(z_k - H\hat{X}_k^-). \quad (3.12)$$

A diferença $(z_k - H\hat{X}_k^-)$ na Equação (3.12) é chamada de “diferença residual”. A diferença residual reflete a discrepância entre a predição da medição $H\hat{X}_k^-$ e a medição atual z_k . Uma diferença residual de média zero significa que os dois estão em completo acordo. A matriz K da Equação (3.12) é escolhida para ser o ganho que minimiza a Equação (3.11) de covariância do erro *a posteriori*. Como todas as equações do filtro de Kalman podem ser manipuladas algebricamente, um modelo do ganho de Kalman K que minimiza a Equação (3.11) é dado por:

$$K_k = P_k^- H^T (H P_k^- H^T + R)^{-1} = \frac{P_k^- H^T}{H P_k^- H^T + R}. \quad (3.13)$$

Observando a Equação (3.13), nota-se que quando a covariância do ruído de medição R se aproxima de zero, o ganho K pondera a diferença residual mais intensamente. Por outro lado, quando a covariância

do erro estimado *a priori* P_k^- se aproxima de zero, o ganho K pondera a diferença residual menos intensamente. Assim:

$$\lim_{R_k \rightarrow 0} K_k = H^{-1}, \quad (3.14)$$

$$\lim_{P_k^- \rightarrow 0} K_k = 0. \quad (3.15)$$

Outra maneira de pensar sobre as ponderações realizadas pelo ganho de Kalman K é que conforme a covariância do ruído de medição R se aproxima de zero, a medição atual z_k seria mais e mais “confiável”, enquanto que a predição da medição $H\hat{X}_k^-$ se tornaria cada vez menos confiável. Por outro lado, conforme a covariância do erro estimado *a priori* P_k^- se aproximasse de zero, a medição atual z_k se tornaria cada vez menos confiável, enquanto que a predição da medição $H\hat{X}_k^-$ se tornaria mais e mais confiável.

3.3.3 Origem Probabilística do Filtro de Kalman

A justificativa para a Equação (3.12) está enraizada na probabilidade da estimativa *a priori* \hat{X}_k^- condicionada a todas as medições z_k anteriores (regra de Bayes). Assim:

$$\begin{aligned} E[X_k] &= \hat{X}_k, \\ E[(X_k - \hat{X}_k)(X_k - \hat{X}_k)^T] &= P_k. \end{aligned}$$

O estado estimado *a posteriori* da Equação (3.12) reflete a média da distribuição do estado. A covariância do erro estimado *a posteriori* da Equação (3.11) reflete a variância da distribuição do estado. Em outras palavras,

$$\begin{aligned} p(X_k | z_k) &\sim N(E[X_k], E[(X_k - \hat{X}_k)(X_k - \hat{X}_k)^T]) \\ &= N(\hat{X}_k, P_k). \end{aligned}$$

3.3.4 Algoritmo do Filtro de Kalman

O filtro de Kalman estima um processo utilizando uma forma de controle por realimentação. O filtro estima o estado do processo em algum momento e então obtêm a realimentação na forma de medições

ruidosas. Além disso, as equações do filtro de Kalman se resumem em dois grupos: **equações de atualização de tempo** e **equações de atualização da medição**. As equações de atualização de tempo são responsáveis por projetar para a frente no tempo o estado atual e covariância do erro estimados, de forma a obter a estimativa *a priori* para o próximo instante de tempo. Já as equações de atualização da medição são responsáveis pela realimentação, por exemplo, incorporar uma nova medição na estimativa *a priori* de forma a obter uma estimativa *a posteriori* melhorada.

As equações de atualização de tempo também podem ser entendidas como **equações de predição**, enquanto que as equações de atualização da medição podem ser compreendidas como **equações de correção**. De fato, o algoritmo de estimação remonta ao algoritmo preditor-corretor com o objetivo de resolver problemas numéricos, como pode ser visto na Figura 5.

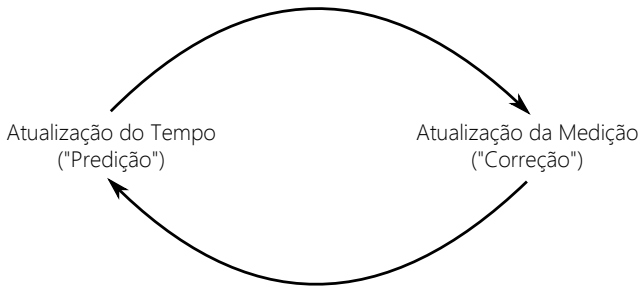


Figura 5 – O ciclo do filtro de Kalman. A **atualização de tempo** projeta a próxima estimativa do estado atual. A **atualização de medição** ajusta a estimativa projetada através de medição real naquele tempo. Baseado em Welch e Bishop (2001).

As equações específicas de predição são dadas por:

$$\hat{X}_k^- = A\hat{X}_{k-1} + Bu_{k-1}, \quad (3.16)$$

$$P_k^- = AP_{k-1}A^T + Q. \quad (3.17)$$

A Equação (3.16) realiza a estimação do estado atual, enquanto que a Equação (3.17) realiza a estimação da covariância do erro, ambas no instante de tempo $k - 1$ até k . Nas equações, A e B referem-se as matrizes da Equação (3.4), Q refere-se a matriz da Equação (3.6).

As equações de correção do filtro de Kalman são dadas por:

$$K_k = P_k^- H^T (H P_k^- H^T + R)^{-1}, \quad (3.18)$$

$$\hat{X}_k = \hat{X}_k^- + K(z_k - H \hat{X}_k^-), \quad (3.19)$$

$$P_k = (I - K_k H) P_k^-. \quad (3.20)$$

Note que as equações (3.18) e (3.13), e (3.19) e (3.12), são as mesmas.

Para inicializar o filtro de Kalman, faz-se necessário a especificação de um estado estimado *a posteriori* inicial \hat{X}_0 e um valor inicial para a covariância do erro estimado *a posteriori* P_0 . Dados os valores para inicialização, a primeira tarefa durante a atualização da medição é computar o ganho de Kalman K_k . O próximo passo é realizar uma medição para obter z_k e gerar uma estimativa do estado *a posteriori* como pode ser visto na Equação (3.19). O passo final é obter uma estimativa *a posteriori* da covariância do erro. A cada atualização de tempo e medição, o processo é repetido com as estimações *a posteriori* anteriores utilizadas para projetar as novas estimativas *a priori*. Esta natureza recursiva é uma característica interessante do filtro de Kalman. A Figura 6 oferece uma visão completa da operação do filtro.

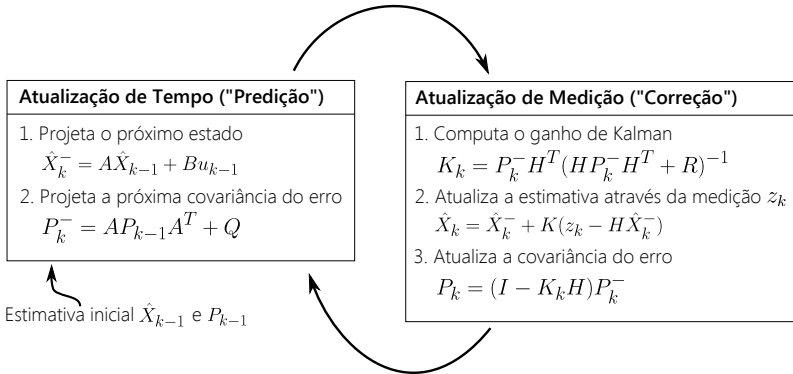


Figura 6 – Operação do filtro de Kalman. Baseado em Welch e Bishop (2001).

3.3.5 Parametrização do Filtro de Kalman

O filtro de Kalman possui dois parâmetros importantes: a covariância do ruído de medição R , e a covariância do ruído do processo Q . Para implementação real do filtro de Kalman, a covariância do ruído de medição R é geralmente definida antes da operação do filtro. Para isso, faz-se necessário obter algumas medidas de amostras *offline* com o objetivo de determinar a variância do ruído de medição. A determinação da covariância do ruído do processo Q é geralmente mais difícil pois tipicamente não é possível observar diretamente o processo a ser estimado. Algumas vezes, um modelo de processo relativamente simples pode produzir resultados aceitáveis se “injeta” incerteza suficiente no processo através da escolha de Q .

3.4 SUMÁRIO

Neste capítulo foram apresentados os conceitos necessários para o entendimento deste trabalho, como o sistema utilizado para localização de veículos e os fundamentos estatísticos empregados. Abordou-se os conceitos sobre o teste de Kolmogorov-Smirnov, uma ferramenta matemática utilizada para o teste de hipóteses sobre a distribuição de uma população de dados. Além disso, foram apresentados os fundamentos relacionados ao filtro de Kalman discreto.

4 ANÁLISE ESTATÍSTICA DE DADOS DE POSIÇÃO E VELOCIDADE DE ÔNIBUS URBANO

Neste capítulo são descritos dois métodos para análise estatística dos dados de medição de posição e velocidade do ônibus, o **método de análise longitudinal** e o **método de análise transversal**, desenvolvidos com o objetivo de extrair informações dos dados históricos/simulados do sistema, relevantes à configuração dos parâmetros do filtro de Kalman. Os métodos são baseados no cálculo de médias e variâncias.

4.1 ANÁLISE LONGITUDINAL DE DADOS DE MEDIÇÃO

O primeiro método desenvolvido para análise de dados é o método de **análise longitudinal**. Este método baseia-se em encontrar a média e desvio padrão de dados históricos disponíveis, sejam de velocidades, sejam de posições do ônibus urbano ao longo do trajeto. Entende-se por posição do ônibus a sua distância atual em relação ao ponto de partida. Todavia, o tratamento dos dados de velocidade é diferente do tratamento empregado aos dados de posição. Assim, considerando as matrizes de dados históricos de velocidade $V_{m \times n}$ e posição $S_{m \times n}$, foram desenvolvidos dois procedimentos distintos.

O objetivo do primeiro procedimento é obter a média e desvio padrão da velocidade histórica do ônibus para um determinado trajeto. Para isso, média e desvio padrão são calculados baseando-se em todos os dados históricos de velocidade disponíveis para o trajeto escolhido. A média e desvio padrão da velocidade histórica, são informações relevantes para a configuração de alguns parâmetros do modelo do sistema, como o desvio padrão do ruído de velocidade e a velocidade média do ônibus durante todo o trajeto.

Considerando a matriz de dados de velocidade $V_{i,j}$ representando a velocidade no tempo $t = t_i$ para um experimento j sendo $t_i \in \{0, 1, \dots, t_f\}$, onde t_f é o tempo final (instantes de amostragem, em minutos), a velocidade média histórica será calculada por:

$$\mu_V = \frac{1}{m \times n} \sum_{j=1}^n \sum_{i=1}^m V_{i,j}, \quad (4.1)$$

e a variância será:

$$\sigma_V^2 = \frac{1}{m \times n} \sum_{j=1}^n \sum_{i=1}^m (V_{i,j} - \mu_V)^2. \quad (4.2)$$

O segundo procedimento tem por objetivo calcular a distância média percorrida pelo ônibus a cada minuto, através dos dados de posição do ônibus ao longo do trajeto. Para isso, considerando a matriz de dados de posição $S_{i,j}$ representando uma posição em $t = t_i$ para um experimento j sendo $t_i \in \{0, 1, \dots, t_f\}$ (instantes de amostragem, em minutos), calcula-se a média por meio de:

$$\mu_S = \frac{1}{n \times (m-1)} \sum_{j=1}^n \sum_{i=1}^{m-1} w_{i,j}, \quad (4.3)$$

onde

$$w_{i,j} = S_{i+1,j} - S_{i,j}. \quad (4.4)$$

A variância será calculada por:

$$\sigma_S^2 = \frac{1}{n \times (m-1)} \sum_{j=1}^n \sum_{i=1}^{m-1} (w_{i,j} - \mu_S)^2. \quad (4.5)$$

De forma a ilustrar este procedimento, na Figura 7 são apresentadas duas curvas resultantes de dois experimentos, portanto $j = 2$, para um mesmo trajeto. Nota-se a aplicação da Equação (4.4) necessária para o cálculo de média, variância e desvio padrão.

Na Figura 8, são apresentados os resultados das subtrações (velocidades) e cálculo da média, obtidos com a aplicação do procedimento para um conjunto de dados de posição. São mostrados os resultados das primeiras 200 subtrações. Através da análise da figura, é possível notar que a velocidade do ônibus a cada minuto varia consideravelmente.

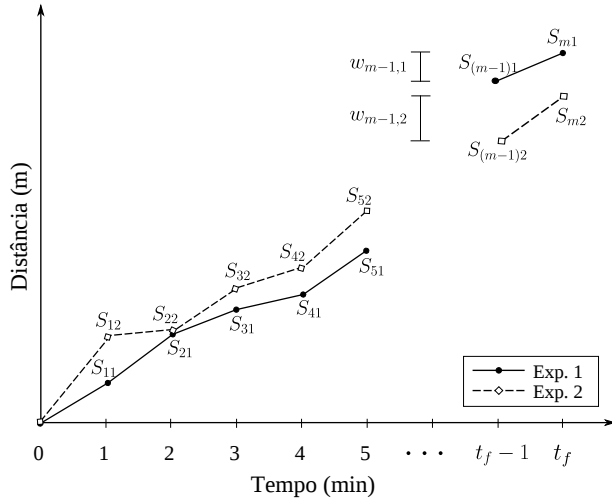


Figura 7 – Representação do método de análise longitudinal para aplicação em dados de posição.

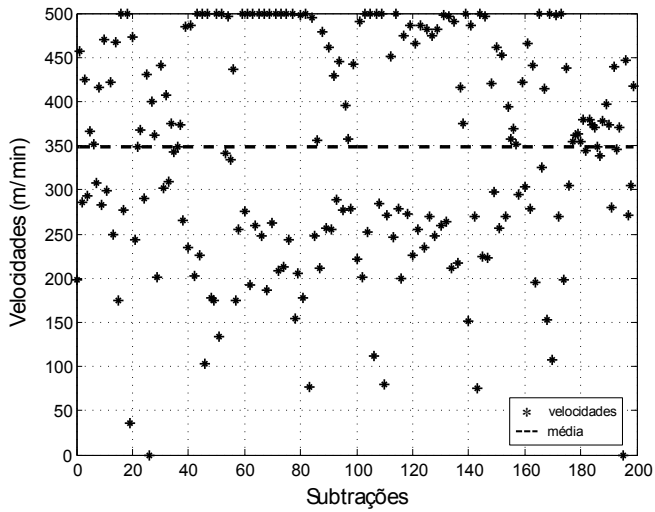


Figura 8 – Subtrações e média resultantes da aplicação do método de análise longitudinal de dados de posição.

A média, a variância e o desvio padrão, obtidos com a implementação dos dois procedimentos, tem resultados iguais. Todavia, o que justifica o desenvolvimento destes é o fato de, em alguns casos, possuir-se somente dados históricos de velocidades ou de posições do veículo para um percurso específico. Para a primeira situação onde há somente dados históricos de velocidades, pode-se obter também média, variância e desvio padrão para posições do ônibus urbano. Caso somente dados históricos de posições estejam disponíveis, é possível obter também média, variância e desvio padrão para velocidades. A aplicação destes dois procedimentos fornece informações relevantes para a implementação do algoritmo do filtro de Kalman proposto neste trabalho.

4.2 ANÁLISE TRANSVERSAL DE DADOS DE MEDIÇÃO

O segundo método desenvolvido para análise dos dados históricos é o método de **análise transversal**.

O principal objetivo deste método é encontrar alterações consideráveis nas variâncias de dados históricos. Para isso, considerando uma matriz de dados de posição $S_{i,j}$, calcula-se média e variância para amostras em t_i para vários experimentos j . Assim:

$$\mu_{S_i} = \frac{1}{n} \sum_{j=1}^n S_{i,j}, \quad (4.6)$$

$$\sigma_{S_i}^2 = \frac{1}{n} \sum_{j=1}^n (S_{i,j} - \mu_{S_i})^2. \quad (4.7)$$

Na Figura 9, é apresentada uma aplicação da análise transversal para três curvas de medidas de posição. As medidas são obtidas em $t_2 = 2$ min, $t_3 = 3$ min, $t_4 = 4$ min e $t_5 = 5$ min. O método calcula a média e variância para as medidas em t_2 , depois para t_3 e assim por diante.

Na Figura 10 podem ser vistas as variâncias obtidas com a análise transversal para um conjunto de dados de posição. É possível observar que a variância muda constantemente, entretanto, segue um comportamento parecido até $t \approx 44$ min, mudando levemente após este instante.

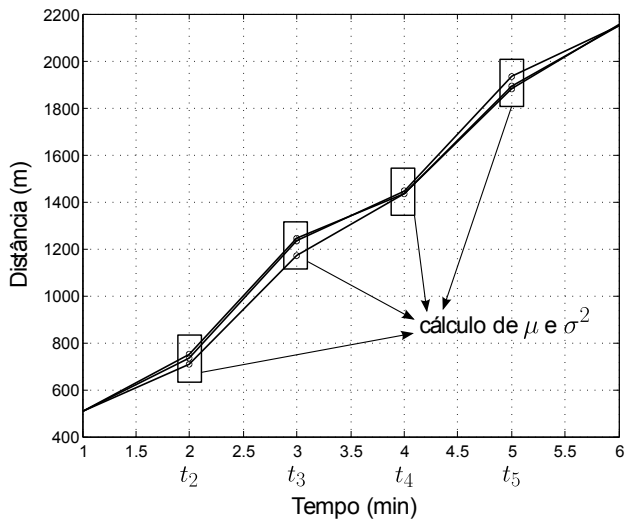


Figura 9 – Representação gráfica do método de análise transversal de dados de posição.

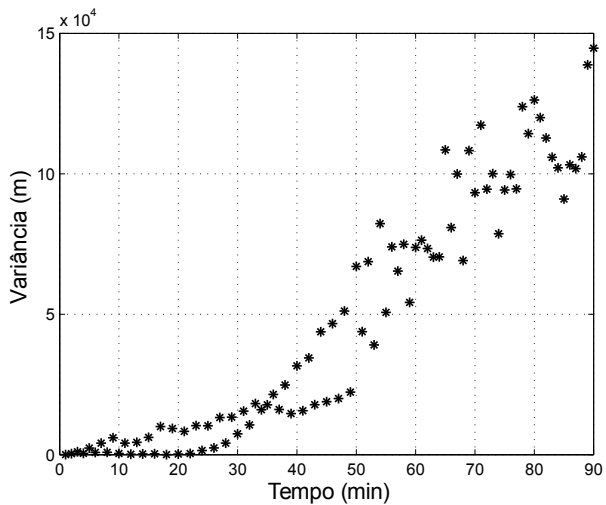


Figura 10 – Variâncias a cada t para dados de posição do ônibus.

O reconhecimento dos pontos onde há mudanças relevantes na variância de dados históricos é importante pois, em alguns casos, a mudança é tão significativa que a distribuição dos dados pode não seguir mais uma Normal, necessária para aplicação do filtro de Kalman. Caso isto aconteça, aplica-se o teste de ajustamento de Kolmogorov-Smirnov no grupo de dados em questão e analisa-se a hipótese destes dados ainda seguirem esta distribuição.

4.3 SUMÁRIO

Neste capítulo, foram exibidos os métodos de análise estatística de dados, desenvolvidos com o objetivo de extrair o maior número possível de informações dos dados históricos do sistema. Estes métodos são importantes para a compreensão da configuração de alguns parâmetros do modelo matemático e do filtro de Kalman do sistema, descritos no próximo capítulo.

5 MODELO DE PREDIÇÃO DE TEMPOS DE CHEGADA DE ÔNIBUS URBANO

Este capítulo apresenta a modelagem matemática do sistema que possibilita, junto ao filtro de Kalman, a estimação de posição, velocidade e tempos de chegada de veículos de transporte coletivo. O estado dinâmico do veículo é estimado através do algoritmo do filtro de Kalman. Os tempos de chegada são estimados através de um algoritmo descrito em sua seção correspondente.

5.1 MODELO PARA ESTIMAÇÃO DE POSIÇÃO E VELOCIDADE DO ÔNIBUS

Neste sistema que envolve ônibus urbano, uma modelagem detalhada da movimentação do veículo é algo inapropriado, e um modelo de movimentação simples é mais adequado (DAILEY et al., 2000). Como dito anteriormente, este trabalho é inspirado nas propostas de Dailey et al. (2000), Wall e Dailey (1999) e principalmente Cathey e Dailey (2003), onde os autores desenvolveram um modelo linear com ruídos para representar a dinâmica do ônibus. Entretanto, o modelo estocástico com aceleração não é necessário para caracterizar o processo de deslocamento do veículo. Portanto, o desenvolvimento realizado por Cathey e Dailey (2003) não será adotado. Apenas os aspectos de modelagem linear do processo serão assumidos.

O filtro de Kalman é aplicado para transformar uma sequência de medições de posição em estimativas dos estados de posição e velocidade do veículo. Entretanto, a fim de implementar o filtro de Kalman, faz-se necessário especificar:

- Um espaço de estados;
- Um modelo de medição;
- Um modelo de transição de estados;
- Um procedimento de inicialização.

Para representar o estado dinâmico instantâneo do veículo, definiu-se um espaço de estados X_k bidimensional. As variáveis de estado no k -ésimo instante são definidas pela posição do veículo x_k e sua velocidade v_k . Assim:

$$X_k = \begin{bmatrix} x_k \\ v_k \end{bmatrix}. \quad (5.1)$$

Uma medição z é uma estimativa da posição do veículo no trajeto. Assim, o modelo de medição é representado por:

$$z_k = HX_k + e_k, \quad (5.2)$$

onde $H = \begin{bmatrix} 1 & 0 \end{bmatrix}$ é a matriz de medição e e_k o erro aleatório de medição. É assumida distribuição normal com média zero para e_k .

A modelagem da dinâmica referente a posição x_k do veículo no trajeto, é baseada nas equações do movimento retilíneo uniforme (MRU). Assim:

$$x_{k+1} = x_k + v_k \Delta t + w_k^x. \quad (5.3)$$

Na Equação (5.3), $x_k + v_k \Delta t$ é a própria função da posição em relação ao tempo do MRU e w_k^x é a parte estocástica assumida como sendo uma distribuição normal com média zero.

Para a modelagem da dinâmica da velocidade do veículo, tem-se:

$$v_{k+1} = v_k + w_k^v, \quad (5.4)$$

onde w_k^v é a parte estocástica que representa a variação de velocidade do veículo, causada por fatores externos como parada em um ponto de ônibus, interseções sinalizadas, entre outros. Assim como o ruído de posição w_k^x , o ruído de velocidade $w_k^v \sim N(0, \sigma^2)$.

Conhecidas estas equações, o problema pode ser colocado na forma do filtro de Kalman linear. Assim, o modelo de transição de estados será:

$$X_{k+1} = AX_k + W_k, \quad (5.5)$$

onde a matriz $n \times n$ A é a matriz de transição de estados, sendo:

$$A = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}. \quad (5.6)$$

A entrada W é o vetor de ruídos do processo. Desta forma:

$$W_k = \begin{bmatrix} w_k^x \\ w_k^v \end{bmatrix}. \quad (5.7)$$

Assim, substituindo as matrizes das Equações (5.1), (5.6) e (5.7)

na Equação (5.5), tem-se:

$$X_{k+1} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ v_k \end{bmatrix} + \begin{bmatrix} w_k^x \\ w_k^v \end{bmatrix}. \quad (5.8)$$

A covariância do ruído de medição R é calculada através do valor esperado do produto entre o erro aleatório de medição e e sua transposta e^T . Todavia, detectou-se através de muitas simulações que, para os dados utilizados neste trabalho, $R \geq 370^2$ obtêm melhores resultados, ou seja, decidiu-se “confiar” mais na predição da medição $H\hat{X}_k^-$ do que na medição atual z_k . Isto é justificado pelo fato que o erro médio ϵ de estimativa de tempo de chegada do ônibus em um ponto de parada qualquer, tem comportamento inversamente proporcional ao erro máximo absoluto E_{\max} de estimação de posição do ônibus ao longo do trajeto. Conforme R se aproxima de zero, o erro E_{\max} se torna menor enquanto que o erro ϵ se torna maior. Como deseja-se um erro ϵ menor, $R \geq 370^2$ mostrou-se mais adequado na obtenção de menores valores para ϵ sem que E_{\max} se tornasse excessivo. Desta forma:

$$R = \begin{cases} 370^2, & \text{se } [e \times e^T] < 370^2 \\ [e \times e^T], & \text{se } [e \times e^T] \geq 370^2 \end{cases}. \quad (5.9)$$

A covariância do ruído do processo Q é obtida através do método disponível em Dailey et al. (2000), onde as variâncias de cada ruído de processo são dispostas na diagonal principal da matriz. Assim:

$$Q = \begin{bmatrix} \sigma_{w_x}^2 & 0 \\ 0 & \sigma_{w_v}^2 \end{bmatrix}. \quad (5.10)$$

O procedimento de inicialização é um método para computar valores iniciais para o vetor de estados e sua matriz de covariância do erro associada. Esses valores iniciais \hat{X}_0 e P_0 são baseados na medição inicial z_0 no tempo t_0 , e na variância da medição R . Assim,

$$\hat{X}_0 = (z_0 \ \psi),$$

$$P_0 = \begin{bmatrix} R & 0 \\ 0 & \sigma_{w_v}^2 \end{bmatrix}.$$

Em \hat{X}_0 , ψ representa uma velocidade estimada baseada em um modelo de velocidade constante (velocidade hipotética):

$$\psi = \frac{X_p}{T_h}, \quad (5.11)$$

onde X_p é a distância máxima do percurso do ônibus e T_h o tempo médio histórico que o veículo demora para percorrer esta distância.

Finalmente, dada uma sequência de medições z_1, z_2, \dots, z_n nos tempos t_1, t_2, \dots, t_n , as equações de transição do filtro de Kalman para este sistema serão:

$$\hat{X}_k^- = A\hat{X}_{k-1}, \quad (5.12)$$

$$P_k^- = AP_{k-1}A^T + Q, \quad (5.13)$$

onde \hat{X}_k^- é o estado estimado *a priori* e \hat{X}_{k-1} é a estimação do estado *a posteriori*.

As equações para atualização de medição do filtro de Kalman para este sistema serão:

$$\hat{X}_k = \hat{X}_k^- + K_k(z_k - H\hat{X}_k^-), \quad (5.14)$$

$$P_k = (I_{2 \times 2} - K_k H)P_k^-. \quad (5.15)$$

A matriz $n \times m$ K é o ganho de Kalman, responsável por minimizar a covariância do erro *a posteriori*. É dado por:

$$K_k = P_k^- H^T (H P_k^- H^T + R)^{-1}. \quad (5.16)$$

Este é o conjunto de equações para as aplicações deste trabalho.

5.2 ALGORITMO PARA PREDIÇÃO DE TEMPOS DE CHEGADA DE ÔNIBUS URBANO

O algoritmo proposto para estimação do tempo de chegada do veículo em um determinado ponto de ônibus, baseia-se nas estimações de posição \hat{x}_k e velocidade \hat{v}_k do ônibus, oriundos do filtro de Kalman; e nas posições dos pontos de ônibus ao longo do trajeto. Entende-se como posição do ponto de ônibus sua distância em relação ao início do trajeto. No Algoritmo 1 são apresentados os passos para estimação de tempo de chegada do veículo em um ponto de ônibus b_i qualquer.

No algoritmo, $BS = [b_1, b_2, \dots, b_N]$ representa um vetor com as posições dos pontos de ônibus presentes no trajeto que o ônibus realizará. As variáveis \hat{x}_k e \hat{v}_k representam as posições e velocidades estimadas através do algoritmo do filtro de Kalman a cada instante de

tempo k . O vetor $\Omega_{BS} = [\Omega_{b_1}, \Omega_{b_2}, \dots, \Omega_{b_N}]$ possui os tempos hipotéticos de chegada do ônibus em cada ponto b_i , calculados por:

$$\Omega_{BS} = \frac{b_i}{\psi}, \quad (5.17)$$

onde b_i é a posição de um ponto de ônibus qualquer e ψ é a velocidade hipotética do ônibus para um trajeto qualquer obtido através da Equação (5.11).

Algoritmo 1. Predição do tempo de chegada de ônibus urbano em um ponto de parada b_i

```

BS = [b1, b2, ..., bN]
ΩBS = [Ωb1, Ωb2, ..., ΩbN], i = 1, ..., N
Υ0bi = Ωbi
for k = 1 to n do
    cálculo de  $\hat{x}_k$  e  $\hat{v}_k$  (filtro de Kalman)
    if  $b_i \geq \hat{x}_k$  then
        |  $\Upsilon_k^{b_i} = \frac{b_i - \hat{x}_k}{\hat{v}_k}$ 
    else
        |  $\Upsilon_k^{b_i} = 0$ 
    end
end

```

O algoritmo é inicializado com a informação das posições dos pontos de ônibus de um trajeto qualquer. Em seguida, informa-se os tempos hipotéticos de chegada do veículo Ω_{b_i} em cada ponto de ônibus b_i deste trajeto. Então, o algoritmo do filtro de Kalman é executado gerando as estimativas de posição e velocidade do veículo. É necessário definir um tempo inicial de chegada do veículo no ponto de ônibus, o tempo inicial considerado será o tempo hipotético de chegada do ônibus no ponto b_i escolhido, portanto, $\Upsilon_0^{b_i} = \Omega_{b_i}$.

Conhecidas estas informações, o algoritmo se baseará em dividir a diferença entre a posição estimada atual do veículo \hat{x}_k e a posição do ponto de ônibus b_i pela velocidade estimada \hat{v}_k . Esta divisão resulta em uma estimativa do tempo de chegada $\Upsilon_k^{b_i}$ do ônibus para cada instante de tempo k .

Se a posição do ponto de ônibus b_i for menor que a posição estimada \hat{x}_k atual do ônibus, o tempo de chegada do veículo no ponto de ônibus será $\Upsilon_k^{b_i} = 0$, pois $b_i < \hat{x}_k$ indica que o ônibus já passou pelo ponto b_i .

5.3 SUMÁRIO

Neste capítulo, apresentou-se a modelagem matemática e a definição dos parâmetros necessários para a aplicação do algoritmo do filtro de Kalman em sistemas de transporte público via ônibus urbano. Foi mostrado também o algoritmo para estimação do tempo de chegada do veículo em um ponto de ônibus qualquer.

6 RESULTADOS NUMÉRICOS E SIMULAÇÕES

No presente capítulo, são apresentados os resultados numéricos obtidos a partir da aplicação dos métodos de análise estatística e algoritmos, propostos neste trabalho, aos dados de medição obtidos com as simulações de dois cenários.

De forma a obter um experimento controlado, foram desenvolvidos dois cenários em que a única variável é o desvio padrão σ do tempo de parada do veículo no ponto de ônibus. O fluxo de veículos na via é zero e os pontos de ônibus são equidistantes. Os cenários transmitem aspectos essenciais do sistema de transporte, quais sejam, tempos de parada estocásticos do veículo nos pontos e incertezas de tempo de viagem. No modelo utilizado, essas últimas são concentradas na variância do tempo de parada do veículo nos pontos de ônibus ao longo do trajeto. Para a simulação dos circuitos, foi utilizado o simulador microscópico de tráfego Aimsun 6.1 (TSS, 2012) em conjunto com uma API (*Application Programming Interface*, ou Interface de Programação de Aplicativos) em Python 2.6 para aquisição dos dados do sistema. Considerou-se distribuição normal para geração dos dados de ambos os cenários.

O teste estatístico de Kolmogorov-Smirnov é aplicado aos dados de medição gerados a partir da simulação dos dois cenários com o objetivo de provar a hipótese de que estes dados sigam uma distribuição normal $N(\mu, \sigma^2)$. Em seguida, são definidos os parâmetros fundamentais para a execução do algoritmo do filtro de Kalman para cada um dos cenários, através da aplicação do método de análise longitudinal. Então, as posições e velocidades do ônibus ao longo do trajeto são estimados e o Algoritmo 1 de estimação de tempo de chegada do ônibus é aplicado para dois pontos de parada do primeiro cenário, e três pontos de parada do segundo cenário. Os resultados das estimativas de tempo de chegada são mostrados e analisados.

6.1 ESTIMAÇÃO DOS TEMPOS DE CHEGADA DO ÔNIBUS PARA O CENÁRIO I

Os dados históricos de posição do ônibus foram gerados através de simulações realizadas no *software* Aimsun 6.1. O resultado da simulação é constituído por duas matrizes de dados, uma de dados de posição $S_{30 \times 20}$ e outra de dados de velocidade $V_{30 \times 20}$.

O cenário é formado por um circuito retangular de 11,2 km com 16 pontos de ônibus equidistantes. As posições dos pontos de parada presentes no trajeto são representadas pelo vetor $BS = [b_1, b_2, \dots, b_N]$, onde b_i representa a posição de um ponto qualquer. A distância entre cada ponto é de 700 m. O circuito completo é ilustrado na Figura 11. O ônibus transita pelo circuito por 30 min, gerando 30 amostras de posição e velocidade, com período de amostragem $\Delta t = 1$ min. Para este cenário, considerou-se velocidade de cruzeiro igual a $V_c = 30$ km/h e média de parada do veículo no ponto de ônibus igual a $\mu = 30$ s. O desvio padrão do tempo de parada do veículo no ponto de ônibus foi mantido em 5 segundos nos primeiros 8 pontos ($\sigma_{b_1, \dots, b_8} = 5$ s), e em 25 segundos nos últimos 8 pontos de ônibus ($\sigma_{b_9, \dots, b_{16}} = 25$ s). O cenário foi simulado 20 vezes.

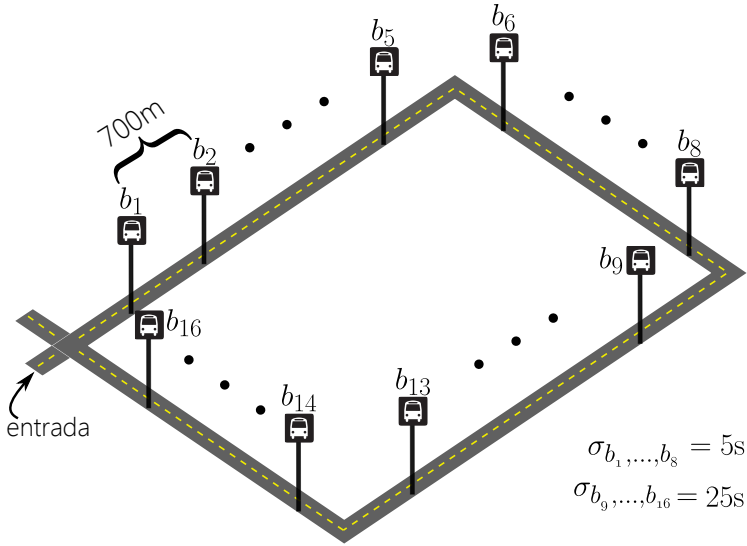


Figura 11 – Cenário de simulação com 16 pontos de parada.

6.1.1 Aplicação do Teste de Kolmogorov-Smirnov

Inicialmente, define-se as hipóteses para aplicação do teste de Kolmogorov-Smirnov. A hipótese nula H_0 é que um conjunto de amostras v_1, v_2, \dots, v_n observadas a partir da população V de amostras de

velocidade do ônibus, siga a distribuição normal $N(\mu, \sigma^2)$. A hipótese alternativa é que estas amostras não sigam esta distribuição. Na forma adimensional, os dados são 292,89; 365,85; 351,57; 307,52; 415,78; 282,52; 470,37; 299,18; 421,51; 249,19.

Baseando-se neste grupo de dados, deseja-se testar a hipótese H_0 de que a distribuição da velocidade segue uma distribuição normal com nível de significância de até 5%. Segundo Soong (2004), este nível de significância $\alpha = 5\%$ é o padrão para o teste de Kolmogorov-Smirnov.

O primeiro passo é organizar os dados de forma crescente. Assim $v_{(1)} = 249,19$; $v_{(3)} = 292,89$; ...; $v_{(10)} = 470,37$. Então, determina-se a função de distribuição $S_n(v)$ das amostras. Portanto,

$$S_n(v) = [0,1; 0,2; \dots; 1].$$

A média $\hat{\mu}$, variância $\hat{\sigma}^2$ e desvio padrão $\hat{\sigma}$ da distribuição teórica $F_V(v)$ são calculadas da seguinte forma:

$$\begin{aligned}\hat{\mu} &= \frac{1}{10} \sum_{i=1}^{10} v_i = 345,63, \\ \hat{\sigma}^2 &= \frac{1}{10} \sum_{i=1}^{10} (v_i - 345,63)^2 = 5149,49, \\ \hat{\sigma} &= \sqrt{5149,49} = 71,76.\end{aligned}$$

Os valores de $F_V(v)$ são encontrados através da Equação (3.3), baseando-se na distribuição $N(\hat{\mu} = 345,63; \hat{\sigma} = 71,76)$, e com auxílio da Tabela A.1. Na Tabela 3 são apresentados os valores da distribuição teórica $F_V(v)$ calculados para cada amostra v .

Finalmente, calcula-se a maior distância entre as distribuições $S_n(v)$ e $F_V(v)$. Para isso

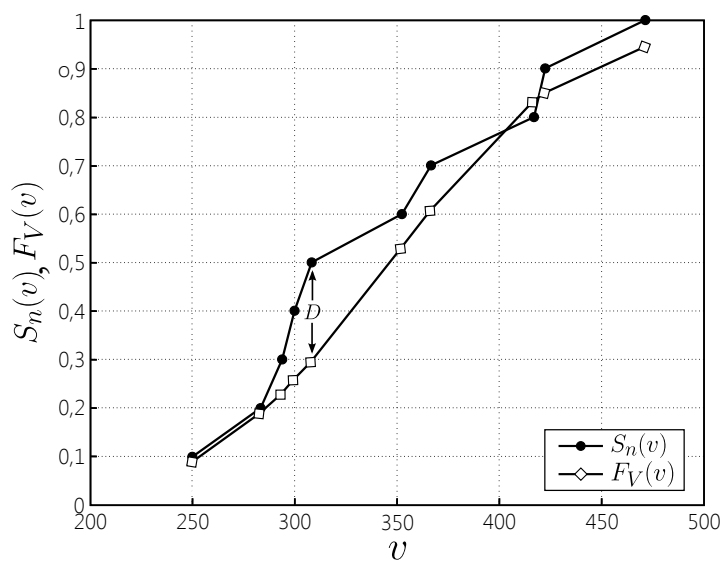
$$D = \max_{i=1}^n |S_n(v) - F_V(v)| = |S_n(307,52) - F_V(307,52)| = 0,2019.$$

Na Figura 12 podem ser vistas as distribuições $S_n(v)$ e $F_V(v)$ plotadas em função de v .

Com $\alpha = 0,05$ e $n = 10$, a Tabela (2) fornece que $d_{0,05}(10) = 0,41$. Portanto, desde que $D < d_{0,05}(10)$, aceita-se a distribuição normal $N(345,63; 71,76)$ com até 5% de nível de significância. Portanto, a hipótese nula H_0 é aceita.

Tabela 3 – Valores da distribuição teórica $F_V(v)$ para cada amostra v .

v	$F_V(v)$
249,19	0,0901
282,52	0,1922
292,89	0,2327
299,18	0,2611
307,52	0,2981
351,57	0,5319
365,85	0,6103
415,78	0,8340
421,51	0,8531
470,37	0,9482

Figura 12 – Resultado do teste de Kolmogorov-Smirnov para os dados do primeiro cenário. Comparação entre $S_n(v)$ e $F_V(v)$.

6.1.2 Configuração do Filtro de Kalman

Após a obtenção e teste dos dados de medição, o primeiro passo é aplicar o método de análise longitudinal descrito no Capítulo 4, desenvolvido com o objetivo de identificar a média μ e desvio padrão σ das velocidades e distâncias percorridas pelo ônibus a cada t . Aplicando o algoritmo, obteve-se:

- $\mu = 348,53$ m
- $\sigma^2 = 14128$ m
- $\sigma = 118,86$ m

Como os dados de posição são iguais aos de velocidades, têm-se:

- $\mu = 348,53$ m/min = 20,9 km/h
- $\sigma^2 = 14128$ m/min = 50,83 km/h
- $\sigma = 118,86$ m/min = 7,13 km/h

Os valores obtidos com a análise longitudinal são úteis na definição do desvio padrão dos ruídos w_x e w_v presentes no sistema. Considerando que os ruídos sigam uma distribuição normal $N(\mu, \sigma)$, têm-se:

- $w_x \sim N(0; 118,86)$
- $w_v \sim N(0; 118,86)$

O desvio padrão do erro de medição e é baseado na precisão do GPS utilizado no ônibus. Supondo que o ônibus esteja utilizando um GPS da marca Garmin modelo eTrex HC, seu erro de precisão será de até 10 metros (GARMIN INTERNATIONAL, INC., 2007). Logo, considerando que o erro aleatório de medição seja uma distribuição normal $N(\mu, \sigma)$, tem-se:

- $e \sim N(0; 10)$

Conhecido o desvio padrão dos ruídos, define-se então as matrizes de covariância do ruído do processo e do ruído de medição.

A matriz Q de covariância do ruído do processo será

$$Q = \begin{bmatrix} \sigma_{w_x}^2 & 0 \\ 0 & \sigma_{w_v}^2 \end{bmatrix} = \begin{bmatrix} 118,86^2 & 0 \\ 0 & 118,86^2 \end{bmatrix}.$$

A matriz R de covariância do ruído de medição será

$$R = \begin{cases} 370^2, & \text{se } [e \times e^T] < 370^2 \\ [e \times e^T], & \text{se } [e \times e^T] \geq 370^2 \end{cases}.$$

O procedimento de inicialização é dado por

$$\hat{X}_0 = (z_0 \ \psi),$$

$$P_0 = \begin{bmatrix} R & 0 \\ 0 & 118,86^2 \end{bmatrix}.$$

A velocidade hipotética ψ é calculada através da Equação (5.11). A distância total do percurso é $X_p = 11200$ m e o tempo médio que o ônibus demora para alcançar esta distância é $T_h = 33$ min. Assim

$$\psi = 339,4 \text{ m/min} = 20,3 \text{ km/h}.$$

6.1.3 Tempos de Chegada Estimados

As posições dos pontos de ônibus no trajeto são representadas pelo vetor $BS = [b_1, b_2, \dots, b_N]$. Para aplicação do algoritmo de tempo de chegada do ônibus urbano no primeiro cenário, escolheu-se os pontos das posições $b_8 = 5800$ m e $b_{14} = 10200$ m. Assim, $BS = [b_8, b_{14}]$.

O passo inicial para estimar os tempos de chegada do ônibus é executar o algoritmo do filtro de Kalman para uma sequência de medições z_1, z_2, \dots, z_{30} nos tempos t_1, t_2, \dots, t_{30} . Todas as simulações para o filtro de Kalman e algoritmo de estimação do tempo de chegada do ônibus, foram realizadas no *software* MATLAB R2010a versão 7.10 de 32 bits.

Na Figura 13 pode ser vista a trajetória estimada do ônibus entre as estimações \hat{x}_0 e \hat{x}_{15} em relação às posições observadas x_0 à x_{15} . As velocidades estimadas podem ser vistas na Figura 15.

Analisando a Figura 13, é possível observar que o algoritmo do filtro de Kalman é eficiente na estimação da posição do ônibus ao longo do trajeto. Todavia, é possível notar erros de estimação em relação as posições observadas. Na Figura 14, são mostrados os erros de estimação da posição do ônibus ao longo do trajeto em relação as posições observadas sem ruído (x_k).

Nota-se que o erro de estimação da posição do ônibus no trajeto, apresentado na Figura 14, cresce até certo ponto a partir de $t = 12$ min.

Isto acontece pelo fato de justamente neste momento o ônibus alcançar o primeiro ponto de parada com desvio padrão igual a 25 segundos.

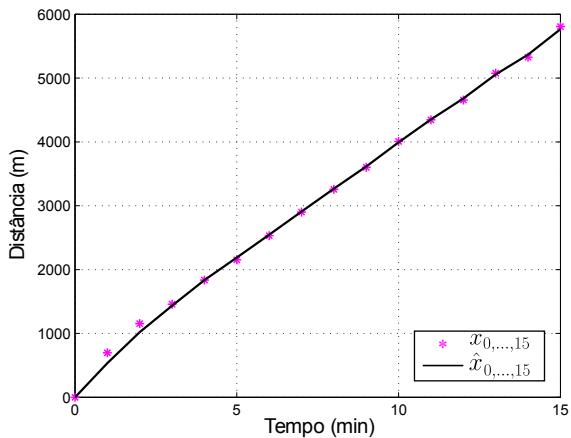


Figura 13 – Estimação da posição do ônibus até a 15ª amostra.

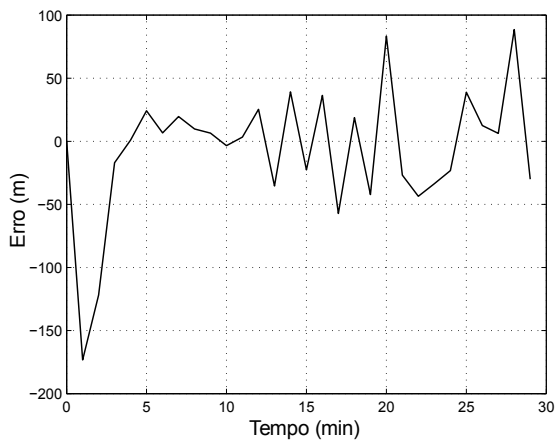


Figura 14 – Erros de estimativa das posições do ônibus ao longo do trajeto em relação às posições observadas sem ruído.

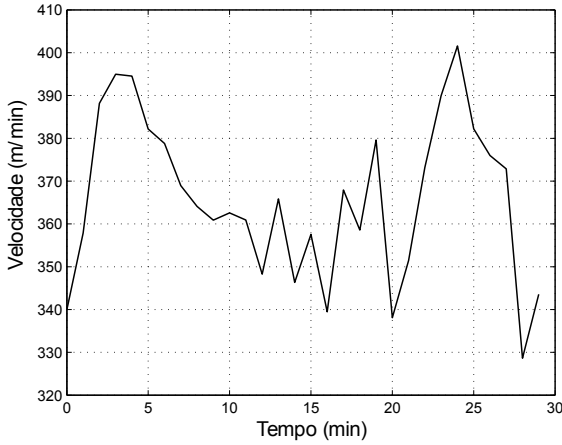


Figura 15 – Velocidades estimadas para os dados de medição do primeiro cenário.

Conforme ocorriam as simulações dos cenários, observou-se que a matriz de covariância do erro estimado *a posteriori* convergia para

$$P = \begin{bmatrix} 79595 & 28453 \\ 28453 & 39521 \end{bmatrix}.$$

Este novo valor de P pode ser aplicado para novas simulações deste cenário, melhorando as estimativas de posição, velocidade e, por consequência, tempos de chegada do ônibus em pontos de parada.

Um dado importante para o operador da linha de ônibus é o erro máximo absoluto E_{\max} da estimativa da posição do veículo no trajeto em relação a posição real do mesmo. Para esta simulação, o erro máximo absoluto encontrado foi

$$E_{\max} \approx 165 \text{ m.}$$

Antes de executar o algoritmo de estimação de tempo de chegada é necessário definir o vetor $\Omega_{BS} = [\Omega_{b_8}, \Omega_{b_{14}}]$ de tempos hipotéticos de chegada do ônibus nos pontos de parada. Aplicando a Equação (5.17) para os pontos de parada em questão, e considerando a velocidade hipotética $\psi = 339,4 \text{ m/min}$, têm-se

$$\Omega_{b_8} = \frac{b_8}{\psi} \approx 17 \text{ min},$$

$$\Omega_{b_{14}} = \frac{b_{14}}{\psi} \approx 30 \text{ min}.$$

Então, aplica-se o Algoritmo 1 para predição do tempo de chegada do ônibus ao ponto de parada b_8 . O tempo de chegada estimado inicial será $\Upsilon_0^{b_8} = \Omega_{b_8}$. O tempo efetivo em que o ônibus chegou ao ponto de parada é $\Gamma_{b_8} = 15\text{min}32\text{s}$. O tempo Γ_{b_8} é conhecido e provém de um banco de dados alimentado pelas simulações do cenário da Figura 11 disponível no Capítulo 4. Na Figura 16 é realizada uma comparação *offline* entre o tempo efetivo Γ_{b_8} de chegada do ônibus ao ponto de parada, o tempo hipotético Ω_{b_8} , e a estimativa do tempo de chegada Υ_{b_8} .

Analisando a Figura 16, observa-se que a curva referente a estimativa do tempo de chegada Υ_{b_8} converge para a reta de referência do tempo efetivo Γ_{b_8} em que o ônibus chegou ao ponto de parada, tendendo a obter um erro de estimativa igual a zero.

Na Figura 17 é apresentado o erro de estimação do tempo de chegada do ônibus em relação ao tempo efetivo Γ_{b_8} em que o mesmo alcançou o ponto de parada b_8 . Nota-se a diminuição do erro de estimação do tempo de chegada conforme o ônibus se aproxima do ponto. Para esta simulação, o erro médio ϵ_{b_8} da estimativa do tempo de chegada em relação ao tempo efetivo foi

$$\epsilon_{b_8} = 23 \text{ s}.$$

O Algoritmo 1 é então aplicado ao ponto de ônibus na posição b_{14} . O tempo de chegada estimado inicial será $\Upsilon_0^{b_{14}} = \Omega_{b_{14}}$. O tempo efetivo em que o ônibus chegou ao ponto de parada é $\Gamma_{b_{14}} = 27\text{min}12\text{s}$. Na Figura 18 é realizada a comparação *offline* entre o tempo efetivo de chegada $\Gamma_{b_{14}}$, o tempo hipotético $\Omega_{b_{14}}$, e a estimativa do tempo de chegada $\Upsilon_{b_{14}}$.

Analisando a Figura 18, observa-se que a curva referente a estimativa do tempo de chegada $\Upsilon_{b_{14}}$ converge para a reta do tempo efetivo de chegada $\Gamma_{b_{14}}$ do ônibus ao ponto de parada, tendendo a obter um erro de estimativa igual a zero. Entretanto, também é possível notar uma alteração no comportamento da curva a partir do tempo $t = 12$ min. Novamente, isto acontece pelo fato de, justamente neste tempo, o ônibus alcançar o primeiro ponto com desvio padrão de tempo de parada $\sigma = 25$ s. O mesmo princípio é aplicado para as curvas de erro

de estimação.

Na Figura 19 é apresentado o erro de estimação do tempo de chegada do ônibus em relação ao tempo efetivo em que o mesmo alcançou o ponto de parada b_{14} . Novamente, percebe-se a diminuição do erro de estimação do tempo de chegada conforme o veículo se aproxima do ponto em questão. Para esta simulação, o erro médio $\epsilon_{b_{14}}$ da estimativa do tempo de chegada em relação ao tempo efetivo foi

$$\epsilon_{b_{14}} \approx 37 \text{ s.}$$

A estimação do tempo de chegada do ônibus para os dois pontos de parada citados se demonstra eficiente. Na Tabela 4 são apresentados os erros médios percentuais ξ_{b_i} para o erro médio ϵ_{b_i} da estimativa do tempo de chegada em relação ao tempo efetivo Γ_{b_i} de chegada do veículo em cada um dos pontos de parada.

Tabela 4 – Erro percentual ξ_{b_i} para o erro médio ϵ_{b_i} da estimativa do tempo de chegada em relação ao tempo efetivo Γ_{b_i} .

BS	Γ_{b_i}	ϵ_{b_i}	ξ_{b_i}
b_8	15min32s	23 s	2,44%
b_{14}	27min12s	37 s	2,25%

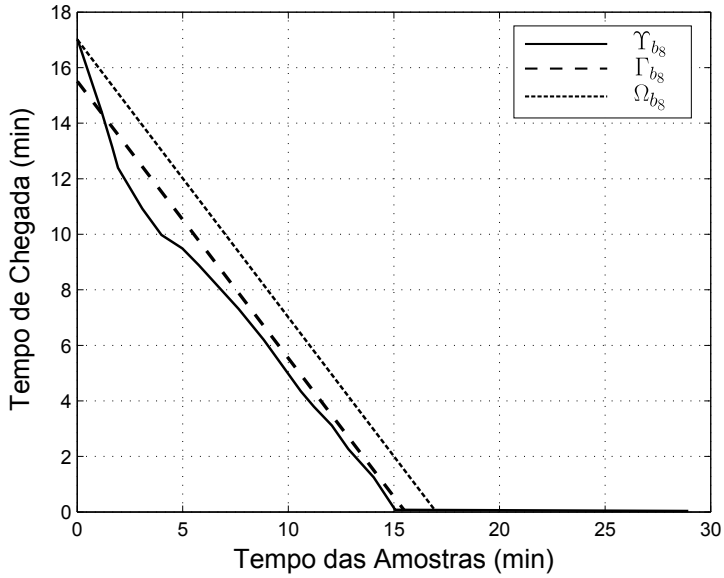


Figura 16 – Tempo efetivo Γ_{bs} , tempo hipotético Ω_{bs} e tempo estimado Υ_{bs} de chegada do veículo ao ponto de parada.

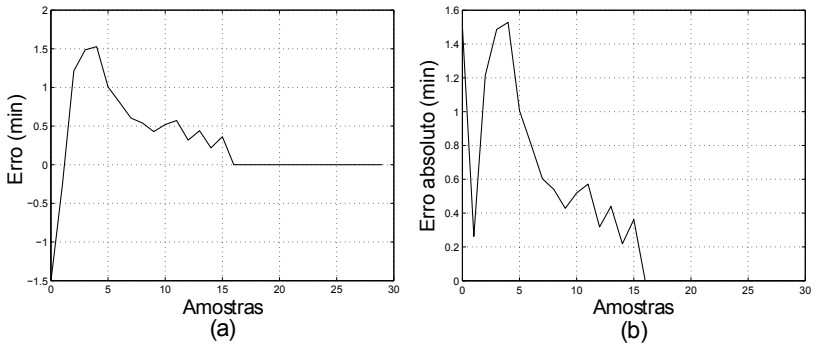


Figura 17 – Erro de estimação do tempo de chegada do ônibus ao ponto de parada b_8 . (a) Erro de estimação; (b) Erro absoluto de estimação.

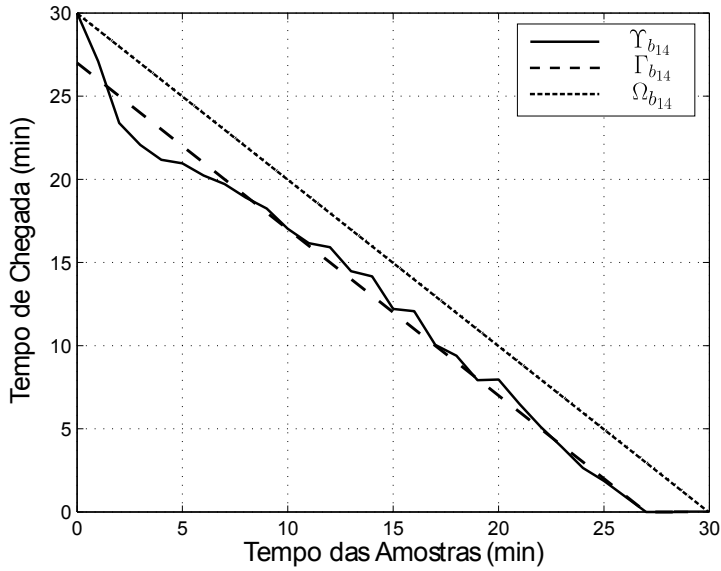


Figura 18 – Tempo efetivo $\Gamma_{b_{14}}$, tempo hipotético $\Omega_{b_{14}}$ e tempo estimado $\Upsilon_{b_{14}}$ de chegada do veículo ao ponto de parada.

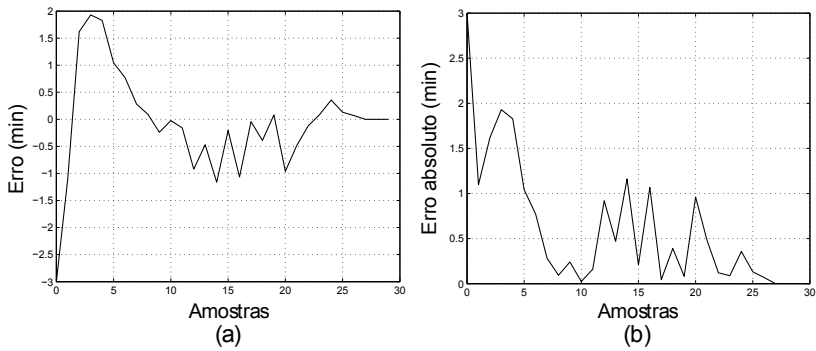


Figura 19 – Erro de estimação do tempo de chegada do ônibus ao ponto de parada b_{14} . (a) Erro de estimação; (b) Erro absoluto de estimação.

Na Figura 20, pode ser visto o erro de estimação dos tempos de chegada do ônibus aos pontos de parada $b_3, b_4, b_5, \dots, b_{16}$ em relação ao ponto b_2 . É percebida uma tendência do erro em se manter próximo a zero, mesmo com o aumento do horizonte de estimação. O erro foi calculado considerando o tempo estimado menos o tempo efetivo de chegada do ônibus aos pontos de parada.

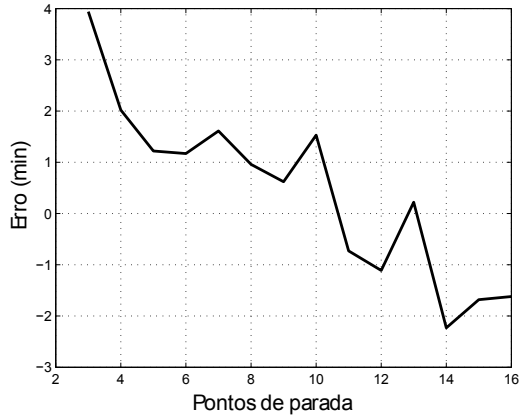


Figura 20 – Erro de estimação dos tempos de chegada do ônibus aos pontos de parada $b_3, b_4, b_5, \dots, b_{16}$ em relação ao ponto b_2 .

O vetor $E = [\zeta_1, \zeta_2, \dots, \zeta_{15}]$ é constituído dos erros de estimação ζ_k dos tempos de chegada do ônibus a cada dois pontos de parada, ou seja, para cada ponto de parada b_i do trajeto, calculou-se o erro de estimação do tempo de chegada ao próximo ponto b_{i+1} . Os erros foram calculados considerando os tempos estimados menos os tempos efetivos de chegada do ônibus aos pontos de parada. Os valores encontrados para E , na forma adimensional, foram: 4,45; 3,94; 2,02; 1,22; 1,17; 1,61; 0,96; 0,62; 1,53; -0,73; -1,12; 0,22; -1,59; -0,64; -0,26.

O teste de Kolmogorov-Smirnov foi aplicado ao vetor E com o objetivo de provar a hipótese de que os erros ζ_k seguem uma distribuição normal.

A média e desvio padrão encontrados, calculados a partir dos dados de E , desconsiderando o período de inicialização do filtro de Kalman, foram:

$$\mu = 0,2,$$

$$\sigma = 1.$$

A estatística de Kolmogorov-Smirnov, resultante da aplicação do teste, foi:

$$D = \max_{i=1}^n |S_n(\zeta) - F_E(\zeta)| = |S_n(0,8) - F_E(0,6591)| = 0,1409.$$

Na Figura 21 podem ser vistas as distribuições $S_n(\zeta)$ e $F_E(\zeta)$ plotadas em função de ζ .

Considerando $\alpha = 0,05$ e $n = 15$, a Tabela 2 fornece um valor crítico de $d_{0,05}(15) = 0,338$. Portanto, desde que $D < 0,338$, aceita-se a distribuição dos dados contidos em E como sendo Gaussiana $N(0,2; 1)$ com até 5% de nível de significância.

A média e desvio padrão (em unidades de tempo) encontrados para os erros de estimação dos tempos de chegada entre pontos de parada, foi:

$$\mu = 0,2 \approx 14 \text{ s},$$

$$\sigma = 1 \approx 1 \text{ min}.$$

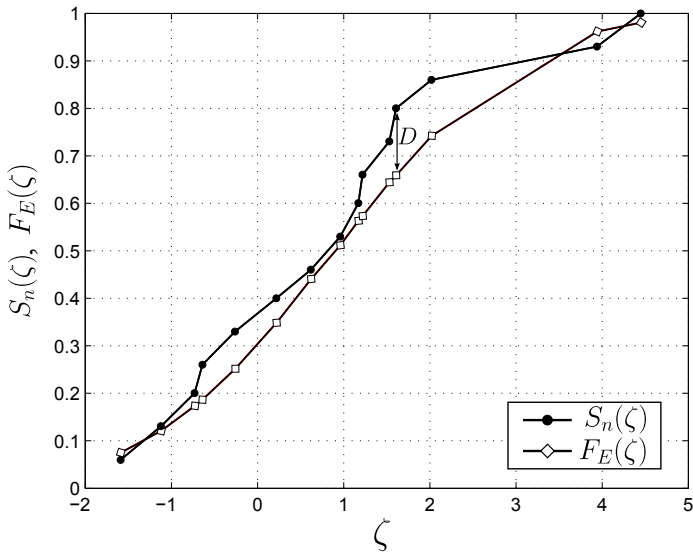


Figura 21 – Resultado do teste de Kolmogorov-Smirnov para os dados do vetor E . Comparação entre $S_n(\zeta)$ e $F_E(\zeta)$.

6.2 ESTIMAÇÃO DOS TEMPOS DE CHEGADA DO ÔNIBUS PARA O CENÁRIO II

Como para o primeiro cenário, para este os dados históricos de medição também foram gerados através de simulações no *software* Aimsun 6.1. O resultado da simulação é constituído por uma matriz de dados de posição $S_{90 \times 20}$ e outra de dados de velocidade $V_{90 \times 20}$.

O cenário é constituído de um circuito também retangular, porém com 31 km de extensão e 44 pontos de ônibus equidistantes. A distância entre cada ponto foi mantida em 700 m. O circuito completo é ilustrado na Figura 22. O ônibus transita pelo circuito por 90 min, gerando 90 amostras de posição e velocidade, $\Delta t = 1$ min. Para este cenário, também foi considerada uma velocidade de cruzeiro igual a $V_c = 30$ km/h e média de parada do veículo no ponto de ônibus igual a $\mu = 30$ s. O desvio padrão de tempo de parada do veículo no ponto de ônibus foi mantido em 5 segundos nos primeiros 22 pontos ($\sigma_{b_1, \dots, b_{22}} = 5$ s), e em 25 segundos nos últimos 22 pontos de ônibus ($\sigma_{b_{23}, \dots, b_{44}} = 25$ s). O cenário também foi simulado 20 vezes.

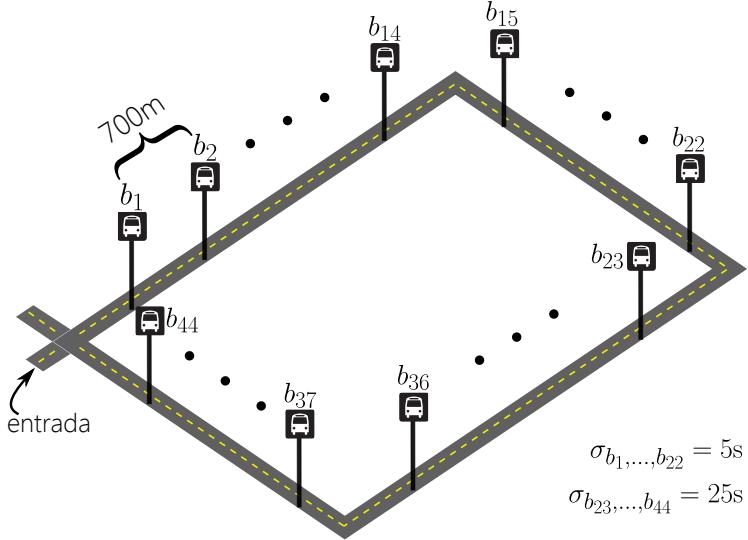


Figura 22 – Cenário de simulação com 44 pontos de parada.

6.2.1 Aplicação do Teste de Kolmogorov-Smirnov

Como para o primeiro cenário, inicialmente define-se a hipótese nula H_0 de que um conjunto de amostras v_1, v_2, \dots, v_n observadas a partir da população V de amostras de velocidade do ônibus, siga a distribuição normal $N(\mu, \sigma^2)$. A hipótese alternativa é que estas amostras não sigam esta distribuição. Na forma adimensional, os dados são 317,20; 394,85; 321,74; 369,85; 347,51; 353,19; 367,98; 328,19; 392,70; 286,52.

Baseando-se neste grupo de dados, deseja-se testar a hipótese H_0 de que a distribuição da velocidade segue uma distribuição normal com nível de significância de até 5%.

Organizando os dados de forma crescente, têm-se

$$v_{(1)} = 286,52; v_{(2)} = 317,20; \dots; v_{(10)} = 394,85.$$

A função de distribuição $S_n(v)$ das amostras será

$$S_n(v) = [0,1; 0,2; \dots; 1].$$

A média $\hat{\mu}$, variância $\hat{\sigma}^2$ e desvio padrão $\hat{\sigma}$ da distribuição teórica $F_V(v)$ serão:

$$\hat{\mu} = \frac{1}{10} \sum_{i=1}^{10} v_i = 347,97,$$

$$\hat{\sigma}^2 = \frac{1}{10} \sum_{i=1}^{10} (v_i - 347,97)^2 = 1211,73,$$

$$\hat{\sigma} = \sqrt{1211,73} = 34,81.$$

Os valores da função de distribuição teórica $F_V(v)$, calculados através da Equação (3.3) com auxílio da Tabela A.1, são baseados na distribuição $N(\hat{\mu} = 347,97; \hat{\sigma} = 34,81)$. Na Tabela 5 são apresentados os valores de $F_V(v)$ para cada amostra v .

Calculando a estatística de Kolmogorov-Smirnov, tem-se

$$D = \max_{i=1}^n |S_n(v) - F_V(v)| = |S_n(328,19) - F_V(328,19)| = 0,1123.$$

Na Figura 23 podem ser vistas as distribuições $S_n(v)$ e $F_V(v)$ plotadas em função de v .

Com $\alpha = 0,05$ e $n = 10$, a Tabela 2 fornece que $d_{0,05}(10) = 0,41$.

Portanto, desde que $D < d_{0,05}(10)$, aceita-se a distribuição normal $N(347,97; 34,81)$ com até 5% de nível de significância. Portanto, hipótese nula H_0 é aceita.

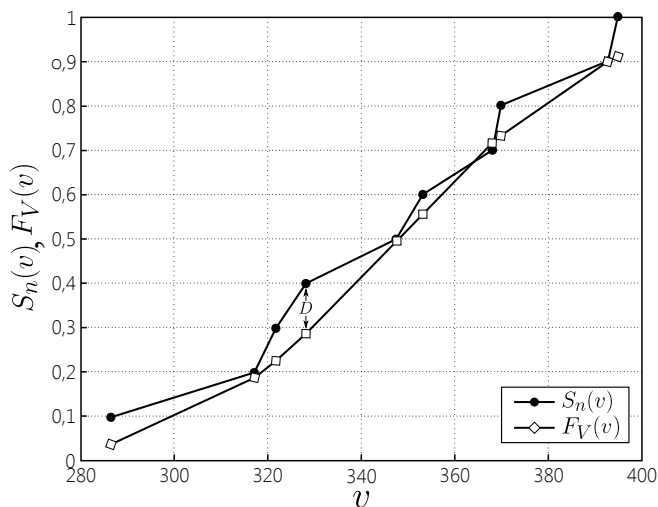


Figura 23 – Resultado do teste de Kolmogorov-Smirnov para os dados do segundo cenário.

Tabela 5 – Valores da distribuição teórica $F_V(v)$ para cada amostra v .

v	$F_V(v)$
286,52	0,0392
317,20	0,1894
321,74	0,2266
328,19	0,2877
347,51	0,4960
353,19	0,5557
367,98	0,7151
369,85	0,7324
392,70	0,8997
394,85	0,9099

6.2.2 Configuração do Filtro de Kalman

Com os dados de medição testados, aplica-se a análise longitudinal. Para este cenário, foram obtidos os seguintes valores referentes às distâncias percorridas a cada t :

- $\mu = 345,81$ m
- $\sigma^2 = 15709$ m
- $\sigma = 125,33$ m

Os valores obtidos referentes às velocidades foram:

- $\mu = 345,81$ m/min = 20,74 km/h
- $\sigma^2 = 15707$ m/min = 56,40 km/h
- $\sigma = 125,33$ m/min = 7,51 km/h

Novamente, considerando a utilização do GPS eTrex HC da Garmin e uma distribuição normal $N(\mu, \sigma)$ para os ruídos, têm-se:

- $w_x \sim N(0; 125,33)$
- $w_v \sim N(0; 125,33)$
- $e \sim N(0; 10)$

A matriz de covariância Q do ruído do processo e a matriz R do ruído de medição serão:

$$Q = \begin{bmatrix} \sigma_{w_x}^2 & 0 \\ 0 & \sigma_{w_v}^2 \end{bmatrix} = \begin{bmatrix} 125,33^2 & 0 \\ 0 & 125,33^2 \end{bmatrix}$$

$$R = \begin{cases} 370^2, & \text{se } [e \times e^T] < 370^2 \\ [e \times e^T], & \text{se } [e \times e^T] \geq 370^2 \end{cases}$$

O procedimento de inicialização também será

$$\hat{X}_0 = (z_0 \ \psi),$$

$$P_0 = \begin{bmatrix} R & 0 \\ 0 & 125,33^2 \end{bmatrix}.$$

A distância total do percurso é $X_p = 31000$ m e o tempo médio que o ônibus demora pra alcançar esta distância é $T_h = 90$ min. Assim, a velocidade hipotética ψ (Equação (5.11)) será

$$\psi = 344,44 \text{ m/min} = 20,66 \text{ km/h.}$$

6.2.3 Tempos de Chegada Estimados

Como no primeiro cenário, as posições dos pontos de ônibus também são representadas no vetor $BS = [b_1, b_2, \dots, b_N]$. Para aplicação do algoritmo de tempo de chegada do ônibus, escolheu-se aleatoriamente os pontos de parada das posições $b_5 = 3590$ m, $b_{25} = 17975$ m e $b_{40} = 28790$ m. Assim, $BS = [b_5, b_{25}, b_{40}]$.

Definidos os pontos de parada, executa-se o algoritmo do filtro de Kalman para um sequência de medições z_1, z_2, \dots, z_{90} para os tempos t_1, t_2, \dots, t_{90} .

Na Figura 24 pode ser vista a trajetória estimada do ônibus entre as estimações \hat{x}_{70} e \hat{x}_{80} em relação às posições observadas sem ruído x_{70} à x_{80} . As velocidades estimadas podem ser vistas na Figura 26.

Analizando a Figura 24, nota-se uma estimacão eficiente da posição do veículo no trajeto. Todavia, é possível observar a existência de erros de estimacão.

Na Figura 25 são apresentados os erros de estimacão da posição do ônibus em relação as posições observadas sem ruído (x_k). Através da análise desta figura, é possível observar uma alteração no comportamento do erro de estimacão de posição à partir de $t = 50$ min. O mesmo acontece com as velocidades, onde também é clara a mudança de comportamento a partir deste tempo. Isto é causado pela alteração do desvio padrão do tempo de parada do ônibus no ponto. Até $t = 50$ min, o desvio padrão do tempo de parada é $\sigma = 5$ seg. Após $t = 50$ min, o desvio padrão é $\sigma = 25$ seg.

Conforme as simulações do cenário ocorriam, observou-se que a matriz de covariância do erro estimado *a posteriori* convergia para

$$P = \begin{bmatrix} 81116 & 29601 \\ 29601 & 43044 \end{bmatrix}.$$

Como para o primeiro cenário, neste, este novo valor de P também pode ser aplicado para novas simulações, melhorando as estimacões de posição, velocidade e, por consequência, tempos de chegada do ônibus em pontos de parada.

O erro máximo absoluto E_{\max} de estimação da posição do veículo no trajeto em relação a sua posição real foi

$$E_{\max} \approx 171,4 \text{ m.}$$

Os tempos hipotéticos de chegada do ônibus nos pontos de parada são calculados através da Equação (5.17). Sabendo que a velocidade hipotética é $\psi = 344,44 \text{ m/min}$, têm-se

$$\Omega_{b_5} = \frac{b_5}{\psi} \approx 10\text{min}25\text{s},$$

$$\Omega_{b_{25}} = \frac{b_{25}}{\psi} \approx 52 \text{ min},$$

$$\Omega_{b_{40}} = \frac{b_{40}}{\psi} \approx 83\text{min}30\text{s}.$$

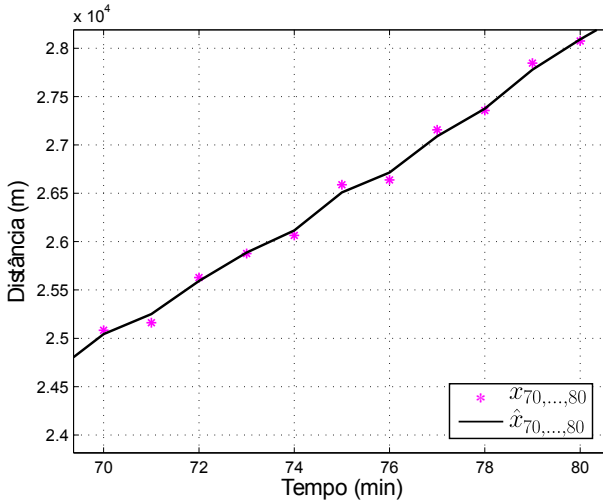


Figura 24 – Estimação de posição do ônibus entre as amostras 70 e 80.

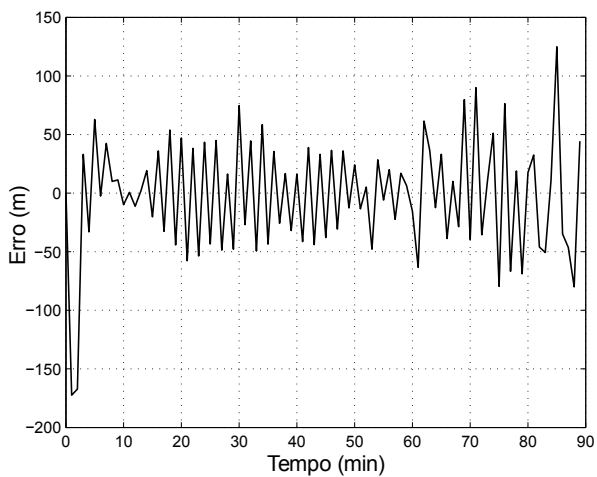


Figura 25 – Erro de estimação da posição do ônibus ao longo do trajeto em relação às posições observadas sem ruído.

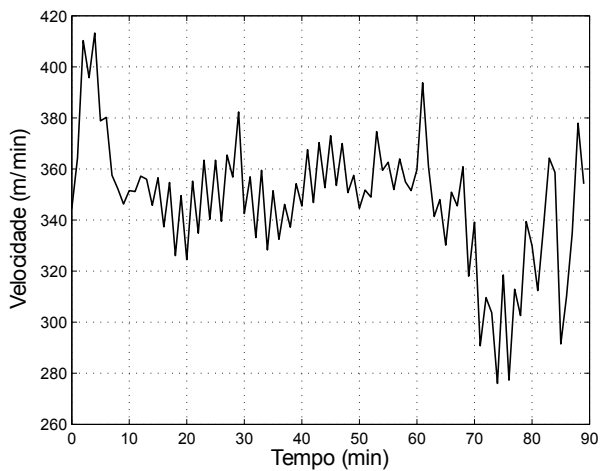


Figura 26 – Velocidades estimadas para os dados de medição do segundo cenário.

Então, o Algoritmo 1 para estimação do tempo de chegada do ônibus é aplicado para o ponto de parada b_5 . O tempo de chegada estimado inicial será $\Upsilon_0^{b_5} = \Omega_{b_5}$. O tempo efetivo em que o ônibus chegou ao ponto de parada foi $\Gamma_{b_5} = 9\text{min}20\text{s}$. Na Figura 27 pode ser vista uma comparação *offline* entre os tempos efetivo Γ_{b_5} , hipotético Ω_{b_5} e estimado Υ_{b_5} de chegada do ônibus ao ponto de parada.

Na Figura 28 é apresentado o erro de estimação do tempo de chegada do ônibus em relação ao tempo efetivo em que o mesmo alcançou o ponto de parada b_5 . Nota-se uma diminuição do erro conforme o ônibus se aproxima do ponto de parada. O erro médio ϵ_{b_5} da estimativa do tempo de chegada em relação ao tempo efetivo foi

$$\epsilon_{b_5} = 33 \text{ s.}$$

Para aplicação do Algoritmo 1 no ponto de parada b_{25} , o tempo de chegada estimado inicial será $\Upsilon_0^{b_{25}} = \Omega_{b_{25}}$, enquanto que o tempo efetivo em que o ônibus chegou ao ponto de parada é $\Gamma_{b_{25}} = 49\text{min}46\text{s}$. Na Figura 29 é realizada uma comparação *offline* entre o tempo efetivo $\Gamma_{b_{25}}$ de chegada do ônibus ao ponto de parada, o tempo hipotético $\Omega_{b_{25}}$, e a estimação do tempo de chegada $\Upsilon_{b_{25}}$.

O erro de estimação do tempo de chegada do ônibus em relação ao tempo efetivo em que o mesmo alcançou o ponto de parada b_{25} pode ser visto na Figura 30. Novamente, é notável a diminuição do erro conforme o ônibus se aproxima do ponto de parada. Para esta simulação, o erro médio $\epsilon_{b_{25}}$ da estimativa do tempo de chegada em relação ao tempo efetivo foi

$$\epsilon_{b_{25}} = 1\text{min}5\text{s.}$$

Finalmente, o Algoritmo 1 é aplicado para o ponto de parada b_{40} . O tempo de chegada estimado inicial será $\Upsilon_0^{b_{40}} = \Omega_{b_{40}}$. O tempo efetivo em que o ônibus chegou ao ponto de parada é $\Gamma_{b_{40}} = 81 \text{ min}$. Novamente, na Figura 31 pode ser vista a comparação *offline* entre os tempos efetivo $\Gamma_{b_{40}}$, hipotético $\Omega_{b_{40}}$ e estimado $\Upsilon_{b_{40}}$ de chegada do ônibus ao ponto de parada.

O erro de estimação do tempo de chegada do ônibus em relação ao tempo efetivo em que o mesmo alcançou o ponto de parada b_{40} pode ser visto na Figura 30. Outra vez, o erro diminui conforme o ônibus se aproxima do ponto de parada. Para esta simulação, o erro médio $\epsilon_{b_{40}}$ da estimativa do tempo de chegada em relação ao tempo efetivo foi

$$\epsilon_{b_{40}} = 2\text{min}5\text{s.}$$

A estimação do tempo de chegada do ônibus para os três pontos de parada citados se demonstra eficiente. Na Tabela 6 são apresentados os erros médios percentuais ξ_{b_i} para o erro médio ϵ_{b_i} da estimativa do tempo de chegada em relação ao tempo efetivo Γ_{b_i} de chegada do veículo em cada um dos pontos de parada.

Tabela 6 – Erro percentual ξ_{b_i} para o erro médio ϵ_{b_i} da estimativa do tempo de chegada em relação ao tempo efetivo Γ_{b_i} .

BS	Γ_{b_i}	ϵ_{b_i}	ξ_{b_i}
b_5	9min20s	33 s	5,89%
b_{25}	49min46s	1min5s	2,17%
b_{40}	81 min	2min5s	2,56%

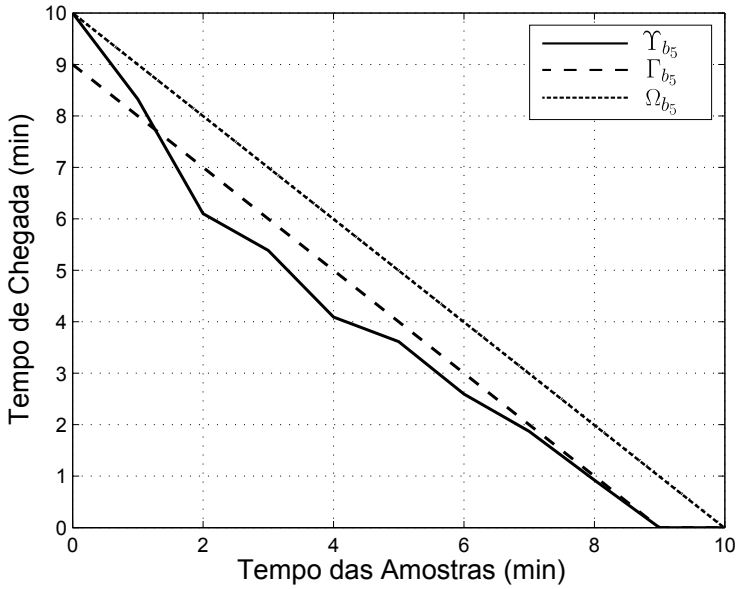


Figura 27 – Tempo efetivo Γ_{b_5} , tempo hipotético Ω_{b_5} e tempo estimado Υ_{b_5} de chegada do veículo ao ponto de parada.

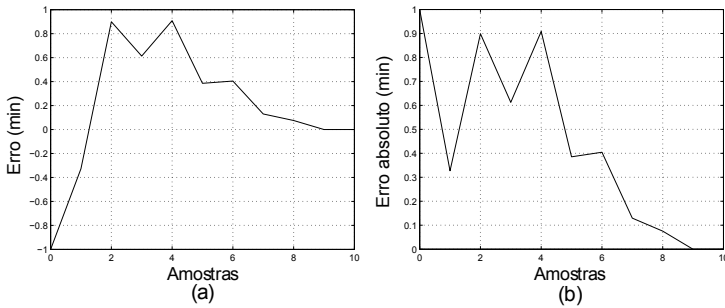


Figura 28 – Erro de estimação do tempo de chegada do ônibus ao ponto de parada b_5 . (a) Erro de estimação; (b) Erro absoluto de estimação.

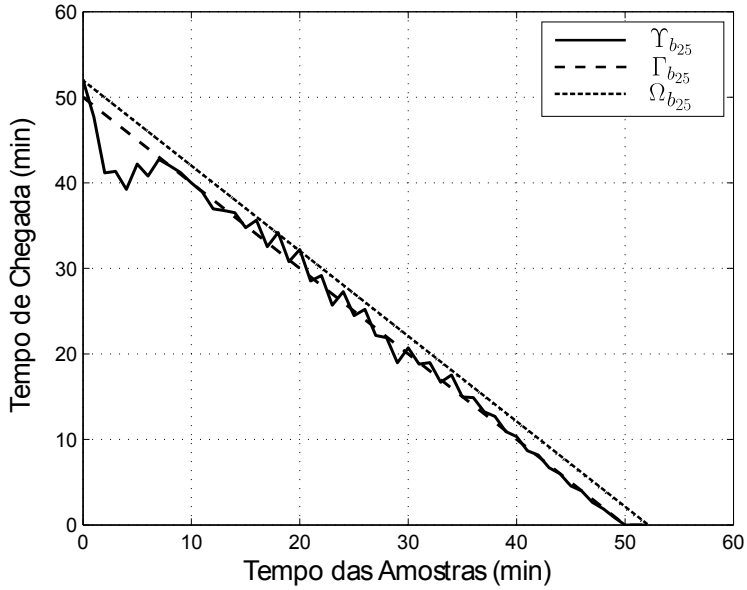


Figura 29 – Tempo efetivo $\Gamma_{b_{25}}$, tempo hipotético $\Omega_{b_{25}}$ e tempo estimado $\Upsilon_{b_{25}}$ de chegada do veículo ao ponto de parada.

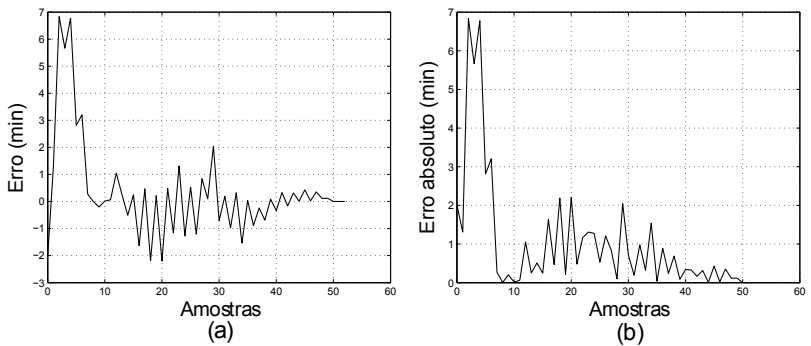


Figura 30 – Erro de estimação do tempo de chegada do ônibus ao ponto de parada b_{25} . (a) Erro de estimação; (b) Erro absoluto de estimação.

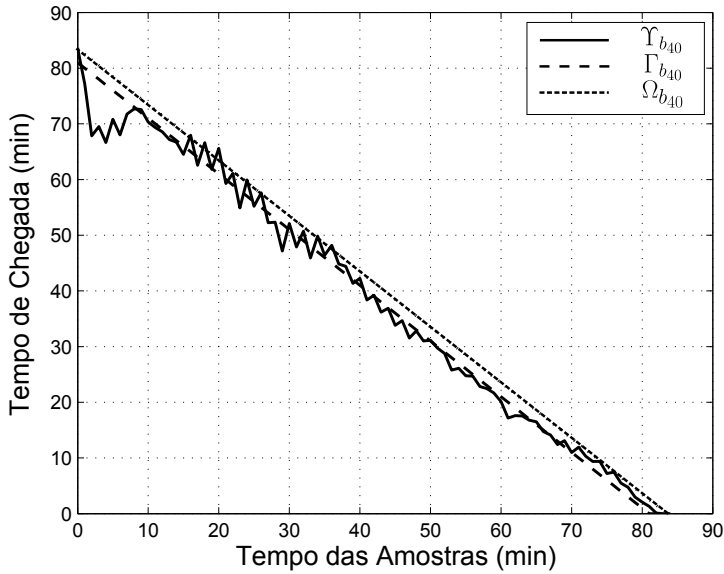


Figura 31 – Tempo efetivo $\Gamma_{b_{40}}$, tempo hipotético $\Omega_{b_{40}}$ e tempo estimado $\Upsilon_{b_{40}}$ de chegada do veículo ao ponto de parada.

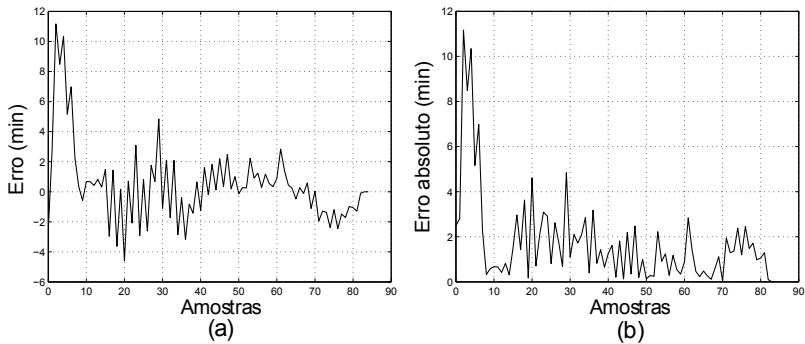


Figura 32 – Erro de estimação do tempo de chegada do ônibus ao ponto de parada b_{40} . (a) Erro de estimação; (b) Erro absoluto de estimação.

6.3 SUMÁRIO

Neste capítulo foram apresentados resultados numéricos para aplicação do algoritmo de estimação do tempo de chegada do ônibus em dois pontos de parada de um cenário de 11,2 km e para três pontos de um cenário de 31 km.

Primeiro, a distribuição estatística dos dados de medição de cada cenário foi avaliada através do teste de Kolmogorov-Smirnov com o objetivo de analisar a hipótese de que estes dados seguissem uma distribuição normal $N(\mu, \sigma^2)$.

Através da aplicação do método de análise longitudinal, os parâmetros necessários para a execução do algoritmo do filtro de Kalman foram definidos. Então, as posições e velocidades do veículo ao longo dos trajetos de cada cenário são estimados através do filtro.

Finalmente, através do algoritmo de estimação do tempo de chegada do ônibus em um ponto de parada, estimou-se os tempos de chegada do veículo em cinco pontos de parada distintos. Pôde-se verificar, através dos gráficos apresentados, que o algoritmo de estimação do tempo de chegada do ônibus em pontos de parada realiza uma estimação eficiente, com erro relativamente pequeno.

7 CONCLUSÃO

Neste trabalho foi proposto um procedimento para estimação de posição, velocidade e tempos de chegada ônibus urbano em pontos de parada, utilizando a abordagem de filtros de Kalman como principal ferramenta de predição. O procedimento consiste em dois métodos para análise estatística dos dados de medição; um modelo linear com ruídos para representação da movimentação do ônibus, baseado no trabalho de Cathey e Dailey (2003); e um algoritmo para predição dos tempos de chegada.

O primeiro método de análise estatística desenvolvido foi o método de análise longitudinal. Este método, teve o objetivo de extrair informações dos dados históricos de posição e velocidade do ônibus, possibilitando um ajuste fino dos parâmetros do filtro de Kalman, de forma a garantir boa estimação dos tempos de chegada em pontos de parada. O segundo, o método de análise transversal, possibilitou uma visão geral dos dados históricos de posição de um determinado trajeto, tornando possível a observação da necessidade de se aplicar um teste de ajustamento, para que fosse identificado se a distribuição destes dados seguiam uma Distribuição Normal, necessária para a aplicação do algoritmo do filtro de Kalman.

O teste de ajustamento de Kolmogorov-Smirnov foi o escolhido, aplicado com o objetivo de avaliar a hipótese de que os dados de medição seguiam uma distribuição normal $N(\mu, \sigma^2)$. Os resultados demonstraram que os dados seguem esta distribuição com nível de significância de até 5%, aceitando, portanto, a hipótese em questão.

Para representação da movimentação do ônibus, utilizou-se de um modelo linear com ruídos, baseado no trabalho de Cathey e Dailey (2003). Este modelo, junto a aplicação da abordagem de filtros de Kalman, foi utilizado para estimação de posição e velocidade do ônibus ao longo dos trajetos dos dois cenários desenvolvidos. Com a configuração correta dos parâmetros através da aplicação do método de análise longitudinal, o modelo se mostrou eficiente, estimando posição e velocidade de maneira satisfatória, contribuindo para a diminuição dos erros de predição dos tempos de chegada do veículo aos pontos de parada escolhidos.

Para predição dos tempos de chegada, foi idealizado um algoritmo que se baseia nas estimativas de posição e velocidade do ônibus, oriundas do algoritmo do filtro de Kalman. A validade deste algoritmo foi mostrada através da aplicação do mesmo para predição dos tempos

de chegada do ônibus em cinco pontos de parada, presentes nos dois cenários de simulação desenvolvidos. Os resultados foram satisfatórios quanto a predição dos tempos de chegada, obtendo erro percentual médio menor que 3%.

Como validação do procedimento desenvolvido para predição de tempos de chegada de ônibus urbano em pontos de parada, foi publicado um artigo no XXVI ANPET - Congresso de Ensino e Pesquisa em Transportes de 2012 (GURGEL; MORENO; KRAUS, 2012).

7.1 TRABALHOS FUTUROS

Propõe-se como trabalhos futuros:

- A aplicação dos métodos de análise estatística de dados, o modelo de filtro de Kalman e o algoritmo de predição de tempos de chegada do ônibus em pontos de parada, a um cenário que contenha interseções semaforizadas, tráfego de veículos e incidentes;
- A aplicação dos métodos de análise estatística de dados, o modelo de filtro de Kalman e o algoritmo de predição de tempos de chegada do ônibus em pontos de parada, a dados de um sistema real de forma a avaliar o desempenho do sistema proposto fora do ambiente simulado.

REFERÊNCIAS

- ABDEL-ATY, M. A. Using ordered probit modeling to study the effect of ATIS on transit ridership. *Transportation Research Part C: Emerging Technologies*, v. 9, n. 4, p. 265–277, 2001.
- ABDELFATTAH, A. M.; KHAN, A. M. Models for predicting bus delays. *Transportation Research Record: Journal of the Transportation Research Board*, v. 1623, p. 8–15, 1998.
- BRANDIMARTE, P. Continuous random variables. Wiley Online Library, 2011.
- BROWN, R. G.; HWANG, P. Y. C. *Introduction to random signals and applied Kalman filtering*. [S.l.]: Wiley New York, 1992.
- CASEY, R. F. et al. *Advanced Public Transportation Systems: The State of the Art*. [S.l.], 1998.
- CATHEY, F. W.; DAILEY, D. J. A prescription for transit arrival/departure prediction using automatic vehicle location data. *Transportation Research Part C: Emerging Technologies*, v. 11, n. 3-4, p. 241–264, 2003.
- CHIEN, S. I.; DING, Y.; WEI, C. Dynamic bus arrival time prediction with artificial neural networks. *Journal of Transportation Engineering*, v. 128, p. 429, 2002.
- CORPUZ, G. Public transport or private vehicle: Factors that impact on mode choice. In: *30th Australasian Transport Research Forum*. [S.l.: s.n.], 2007.
- DAILEY, D. J. et al. An algorithm and implementation to predict the arrival of transit vehicles. *2000 Intelligent Transportation Systems*, p. 161–166, 2000.
- DEFENSE, U. D. of. *User Equipment Introduction*. 1. ed. [S.l.], 1996.
- DELGADO, F. et al. Real-time control of buses in a transit corridor based on vehicle holding and boarding limits. *Transportation Research Record: Journal of the Transportation Research Board*, v. 2090, p. 59–67, 2009.

GARMIN INTERNATIONAL, INC. *Navegador Pessoal Garmin Série eTrex HC - Manual do Proprietário*. [S.l.], Julho 2007.

GUERREIRO, J. C. F. *Fontes de Erros em GPS*. 2005.
<<http://www.malima.com.br>>. Acessado em 01/03/2012.

GURGEL, F. S.; MORENO, U. F.; KRAUS, W. Estimação de tempos de chegada de ônibus urbanos em pontos de parada utilizando filtros de Kalman. *XXVI ANPET - Congresso de Ensino e Pesquisa em Transportes*, 2012.

JAZWINSKI, A. H. *Stochastic processes and filtering theory*. [S.l.]: Academic Press, 1970.

JEONG, R.; RILETT, R. Bus arrival time prediction using artificial neural network model. In: *Intelligent Transportation Systems*. [S.l.: s.n.], 2004. p. 988–993.

JEONG, R. H. *The Prediction of Bus Arrival Time Using Automatic Vehicle Location Systems Data*. Tese (Doutorado) — Texas A&M University, 2004.

JOHN, S. et al. *Global Positioning System: A Guide for the Social Science*. [S.l.]: Blackwell Publishing, 2003.

KALMAN, R. E. A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, v. 82, n. Series D, p. 35–45, 1960.

KALMAN, R. E.; BUCY, R. S. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, v. 83, p. 95, 1961.

KAPLAN, E. D.; HEGARTY, C. J. *Understanding GPS: Principles And Applications*. [S.l.]: Artech House, 2006. (Artech House Mobile Communications Series).

KOEHLER, L. A.; KRAUS, W.; CAMPONOGARA, E. Controle integrado de prioridade e retenção para um sistema BRT. In: *Anais do XXII Congresso de Pesquisa e Ensino em Transportes*. [S.l.: s.n.], 2008.

LAMPARIELLO, F. On the use of the Kolmogorov-Smirnov statistical test for immunofluorescence histogram comparison. *Cytometry*, v. 39, n. 3, p. 179–188, 2000.

LARSON, H. J. *Introduction to probability theory and statistical inference*. [S.l.]: Wiley New York, 1982.

LI, F. et al. Public bus arrival time prediction based on traffic information management system. In: *Service Operations, Logistics, and Informatics - SOLI*. [S.l.: s.n.], 2011. p. 336–341.

LIN, W. H.; ZENG, J. Experimental study of real-time bus arrival time prediction with GPS data. *Transportation Research Record: Journal of the Transportation Research Board*, v. 1666, p. 101–109, 1999.

LIU, H. et al. Predicting urban arterial travel time with state-space neural networks and Kalman filters. *Transportation Research Record: Journal of the Transportation Research Board*, v. 1968, p. 99–108, 2006.

MAJOR, M. Continuous random variable. *Mathematica II*, p. 41, 2009.

MASSEY, F. J. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, v. 46, n. 253, p. 68–78, 1951.

MAYBECK, P. S. *Stochastic models, estimation, and control*. [S.l.: s.n.], 1979. (Mathematics in Science and Engineering, v. 141).

MEINHOLD, R. J.; SINGPURWALLA, N. D. Understanding the Kalman filter. *American Statistician*, p. 123–127, 1983.

MONICO, J. F. G. *Posicionamento pelo NAVSTAR-GPS*. [S.l.]: Unesp, 2000.

MONTGOMERY, D. C.; RUNGER, G. C. *Applied Statistics and Probability for Engineers*. [S.l.]: Wiley, 2010.

NEGENBORN, R. *Robot Localization and Kalman Filters*. Dissertação (Mestrado) — Utrecht University, 2003.

PADMANABAN, R. P. S. et al. Development of a real-time bus arrival prediction system for indian traffic conditions. *Intelligent Transport Systems*, v. 4, n. 3, p. 189–200, 2010.

PRESS, W. H. et al. *Numerical Recipes in C: the art of scientific computing*. [S.l.]: Cambridge University Press, 2002.

- QUDDUS, M. A.; OCHIENG, W. Y.; NOLAND, R. B. Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transportation Research Part C: Emerging Technologies*, v. 15, n. 5, p. 312–328, 2007.
- RISTIC, B.; ARULAMPALAM, S.; GORDON, N. *Beyond the Kalman filter: Particle filters for tracking applications*. [S.l.]: Artech House Publishers, 2004.
- ROSS, S. M. *Introduction to Probability and Statistics for Engineers and Scientists*. [S.l.]: Academic Press, 2004.
- SCHWEIGER, C. L. *Real-time bus arrival information systems*. [S.l.]: Transportation Research Board of The National Academies, 2003.
- SEEMA, S. R.; ALEX, S. Dynamic bus arrival time prediction using GPS data. *10th National Conference on Technological Trends*, Nov 2009.
- SHALABY, A.; FARHAN, A. An integrated modeling system for dynamic operations control and real-time transit information. In: *Intelligent Transportation Systems*. [S.l.: s.n.], 2002. p. 874–879.
- SHILGALIS, T. W. Cumulative distribution functions. *Teaching Statistics*, v. 10, n. 3, p. 82–86, 1988.
- SOONG, T. T. *Fundamentals of probability and statistics for engineers*. [S.l.]: Wiley, 2004.
- STEPHENS, M. A. Use of the kolmogorov-smirnov, cramér-von mises and related statistics without extensive tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 115–122, 1970.
- TAYLOR, G. et al. GPS accuracy estimation using map matching techniques: Applied to vehicle positioning and odometer calibration. *Computers, environment and urban systems*, v. 30, n. 6, p. 757–772, 2006.
- TROSSET, M. *An introduction to statistical inference and its applications with R*. [S.l.]: Chapman & Hall/CRC, 2009.
- TSS. *Traffic Simulation Software*. 2012. <<http://www.aimsun.com>>. Acessado em 12/06/2011.

TURNBULL, K. F. *Evaluation of Automatic Vehicle Location Systems in Public Transit*. [S.l.], 1993.

WALL, Z.; DAILEY, D. J. An algorithm for predicting the arrival time of mass transit vehicles using automatic vehicle location data. *Proceedings of the Transportation Research Board Annual Meeting*, 1999.

WASSERMAN, L. *All of Statistics: a Concise Course in Statistical Inference*. [S.l.]: Springer Verlag, 2004.

WELCH, G.; BISHOP, G. An introduction to the Kalman filter. v. 7, n. 1, p. 1–16, 2001.

ZHU, T. et al. The prediction of bus arrival time using global positioning system data and dynamic traffic information. In: *Wireless and Mobile Networking Conference - WMNC*. [S.l.: s.n.], 2011. p. 1–5.

**APÊNDICE A – Tabela para Funções de Distribuição
Normal Padronizadas**

APÊNDICE B – Probabilidade e Variáveis Aleatórias

B.1 AXIOMAS DE PROBABILIDADE

Wasserman (2004) afirma que probabilidade é uma linguagem matemática para quantificar incertezas. A teoria de probabilidade pode ser aplicada a um conjunto diverso de problemas, desde o lançamento de uma moeda até a análise de algoritmos computacionais complexos.

O modelo matemático que tem dominado o estudo da probabilidade foi formalizado pelo matemático russo A. N. Kolmogorov em uma monografia publicada em 1933. Segundo Trosset (2009), o conceito central neste modelo é o **espaço de probabilidades**, o qual possui três componentes:

1. Um **espaço amostral** \mathbf{S} . Um universo de resultados possíveis para o experimento em questão. Pode ser **discreto** se consiste de um conjunto de resultados finitos ou finitos enumeráveis, e **contínuo** se contém um intervalo (finito ou infinito) de números reais. Por exemplo, lançar uma moeda tem um espaço amostral discreto $\mathbf{S} = \{\textit{cara}, \textit{coroa}\}$.
2. Uma **coleção de eventos** \mathbf{C} . Eventos são subconjuntos do espaço amostral. Por exemplo, supondo que o resultado $\mathbf{s} \in \mathbf{S}$ foi observado e $\mathbf{E} \subset \mathbf{S}$ é um evento, então diz-se que \mathbf{E} ocorrerá se e somente se $\mathbf{s} \in \mathbf{E}$. Quando \mathbf{S} é um conjunto finito, usualmente, todos os subconjuntos de \mathbf{S} são definidos como eventos.
3. Uma **medida de probabilidade** \mathbf{P} . Uma função que associa números reais (chamados probabilidades) à eventos. Uma vez que uma coleção de eventos é definida, cada evento $\mathbf{E} \in \mathbf{C}$ pode ser associado à uma probabilidade $\mathbf{P}(\mathbf{E})$. Isso é realizado através de algumas regras específicas; em particular, a medida de probabilidade \mathbf{P} deve satisfazer:

- Se \mathbf{E} é um evento, então $0 \leq \mathbf{P}(\mathbf{E}) \leq 1$.
- $\mathbf{P}(\mathbf{E}) = 1$.
- Se $\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3, \dots\}$ é uma coleção contável de eventos dois a dois disjuntos, então:

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} \mathbf{E}_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(\mathbf{E}_i) \quad (\text{B.1})$$

A primeira propriedade demonstra que as probabilidades são não-negativas e finitas. Assim, tanto o estado onde “a probabilidade de

que chova hoje é $-0,5$ ” quanto “a probabilidade de que chova hoje é infinita”, não são significativos.

A segunda propriedade expõe que a probabilidade de um resultado ocorrer é no máximo uma unidade. Portanto, o estado “a probabilidade de chover hoje é 2” não é significativo. Esta é uma convenção que simplifica as fórmulas e facilita a interpretação.

A terceira propriedade é chamada de **aditividade enumerável**. Supondo que o lançamento de uma moeda possui os eventos $\{HT\}$ e $\{TH\}$, onde H é *cara* e T representa *coroa*, e que deseja-se computar a probabilidade de que seja observado exatamente um H, a probabilidade de

$$\{HT\} \cup \{TH\} = \{HT, TH\}.$$

Pelo fato de $\{HT\}$ e $\{TH\}$ serem eventos, a união desses resulta em outro evento e portanto tem uma probabilidade. Como são mutuamente exclusivos, ou seja $E_1 \cap E_2 = \emptyset$, deseja-se que a probabilidade seja

$$P(\{HT, TH\}) = P(\{HT\}) + P(\{TH\}).$$

Assim, conclui-se que a probabilidade da união de quaisquer dois eventos disjuntos é a soma de suas respectivas probabilidades.

Assumindo que $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$, torna-se fácil computar a probabilidade de uma união finita de quaisquer eventos dois a dois disjuntos.

Ross (2004), Montgomery e Runger (2010) explicam que dois eventos de um experimento aleatório podem ser independentes. Assim, considerando os eventos A e B , os mesmos serão independentes se $P(AB) = P(A) \times P(B)$ e escreve-se $A \amalg B$. Todavia, a condição de dependência é definida através da teoria de **Probabilidade Condicional**. Assumindo que $P(B) > 0$, a probabilidade condicional de A dada a ocorrência de B será:

$$P(A|B) = \frac{P(B \cap A)}{P(B)}. \quad (B.2)$$

B.2 VARIÁVEIS ALEATÓRIAS DISCRETAS

Segundo Wasserman (2004), a ligação entre espaço amostral e eventos em relação aos dados obtidos através de um experimento aleatório, é realizado através do conceito de **variável aleatória**. Como

o valor de uma variável aleatória é determinado pelo resultado de um experimento, pode-se atribuir probabilidades a seus valores. Formalmente, uma variável aleatória é um mapeamento $\mathbf{X} : \Omega \rightarrow \mathbb{R}$ que atribui um número real $\mathbf{X}(\mathbf{x})$ para cada resultado \mathbf{x} . Por exemplo, lançar uma moeda dez vezes supondo que $\mathbf{X}(\mathbf{x})$ seja o número de vezes que o resultado obtido foi *cara*. Sendo H = *cara* e T = *coroa*, se $\mathbf{x} = \text{HHTHHTHHTT}$, então $\mathbf{X}(\mathbf{x}) = 6$.

Variáveis aleatórias discretas possuem uma extensão finita ou infinita enumerável. São dotadas de uma **distribuição de probabilidade**. Esta distribuição nada mais é que uma descrição das probabilidades associadas com os possíveis valores de uma variável aleatória qualquer \mathbf{X} . No caso de uma variável aleatória discreta, a distribuição é frequentemente especificada apenas por uma lista de valores possíveis, juntamente com a probabilidade de cada um desses valores. Em alguns casos, é conveniente expressar a probabilidade em termos de uma fórmula (MONTGOMERY; RUNGER, 2010).

Shilgalis (1988) e Wasserman (2004) definem uma importante função chamada **função de distribuição acumulada** (ou função de distribuição) que possui todas as informações a respeito de uma variável aleatória. A função de distribuição acumulada (FDA) $\mathbf{F}_{\mathbf{X}} : \mathbb{R} \rightarrow [0, 1]$ de uma variável aleatória \mathbf{X} é definida por

$$\mathbf{F}_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x}). \quad (\text{B.3})$$

Para uma variável aleatória discreta \mathbf{X} , $\mathbf{F}_{\mathbf{X}}(\mathbf{x})$ deve satisfazer as seguintes propriedades:

1. $\mathbf{F}_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x})$;
2. $0 \leq \mathbf{F}_{\mathbf{X}}(\mathbf{x}) \leq 1$;
3. Se $\mathbf{x} \leq \mathbf{y}$, então $\mathbf{F}_{\mathbf{X}}(\mathbf{x}) \leq \mathbf{F}_{\mathbf{X}}(\mathbf{y})$.

Larson (1982) e Wasserman (2004) explicam que uma variável aleatória discreta \mathbf{X} com valores possíveis $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ possui uma **função massa de probabilidade** (FMP) $\mathbf{f}_{\mathbf{X}}(\mathbf{x})$ dada por

$$\mathbf{f}_{\mathbf{X}}(\mathbf{x}) = P\{\mathbf{X} = \mathbf{x}\}.$$

Todavia, $\mathbf{f}_{\mathbf{X}}(\mathbf{x}) \geq 0$ para todo $\mathbf{x} \in \mathbb{R}$ e $\sum_i \mathbf{f}_{\mathbf{X}}(\mathbf{x}_i) = 1$. A FDA de \mathbf{X} é relacionada com $\mathbf{f}_{\mathbf{X}}$ por:

$$\mathbf{F}_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x}) = \sum_{\mathbf{x}_i \leq \mathbf{x}} \mathbf{f}_{\mathbf{X}}(\mathbf{x}_i) \quad (\text{B.4})$$

Dois números são frequentemente utilizados para resumir a distribuição de probabilidade de uma variável aleatória \mathbf{X} . A **média** é uma medida do centro ou meio de uma distribuição de probabilidade, enquanto que a **variância** é uma medida da dispersão ou variação na distribuição. estas duas medidas não identificam exclusivamente uma distribuição de probabilidade, isto é, duas distribuições podem ter a mesma média e variância. Ainda sim, estas medições são simples e úteis para resumir a distribuição de probabilidade de \mathbf{X} (MONTGOMERY; RUNGER, 2010).

Segundo Larson (1982) e Ross (2004), considerando um conjunto de dados com n valores numéricos x_1, x_2, \dots, x_n , a **média** ou **valor esperado** de uma variável aleatória discreta \mathbf{X} , denotada como μ ou $E(\mathbf{X})$, será:

$$\mu = E(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n x_i = \sum_x x f_{\mathbf{X}}(x) \quad (\text{B.5})$$

A **variância** de \mathbf{X} divide a soma dos quadrados das diferenças por $n - 1$ em vez de n , onde n é o tamanho do conjunto de dados. Denotada como σ^2 ou $V(\mathbf{X})$, a variância será:

$$\begin{aligned} \sigma^2 = V(\mathbf{X}) &= E[(\mathbf{X} - \mu)^2] = \frac{1}{n - 1} \sum_{i=1}^n (x_i - \mu)^2 \\ &= \sum_x x^2 f_{\mathbf{X}}(x) - \mu^2 \end{aligned} \quad (\text{B.6})$$

O **desvio padrão** é a raiz quadrada da variância de \mathbf{X} . Portanto $\sigma = \sqrt{\sigma^2}$.

B.3 VARIÁVEIS ALEATÓRIAS CONTÍNUAS

Uma variável aleatória contínua pode assumir qualquer valor dentro de um intervalo (finito ou infinito não enumerável) de números reais (MONTGOMERY; RUNGER, 2010).

Wasserman (2004) e Major (2009) esclarecem que uma variável aleatória \mathbf{X} é **contínua** se existe uma função densidade de probabilidade (FDP) $f_{\mathbf{X}}$ no qual $f_{\mathbf{X}}(x) \geq 0$ para todo x , $\int_{-\infty}^{\infty} f_{\mathbf{X}}(x) dx = 1$ e para cada $a \leq b$,

$$P(a < X < b) = \int_a^b f_X(x)dx. \quad (\text{B.7})$$

Se \mathbf{X} é uma variável aleatória contínua, para quaisquer \mathbf{a} e \mathbf{b} ,

$$\begin{aligned} P(a < X < b) &= P(a \leq X < b) \\ &= P(a < X \leq b) = P(a \leq X \leq b). \end{aligned}$$

A FDP fornece uma descrição simples das probabilidades associadas à uma variável aleatória. Formalmente, uma FDP é uma função $\mathbf{f} : \mathbb{R} \rightarrow \mathbb{R}$ na qual $\mathbf{f}_X(\mathbf{x}) \geq \mathbf{0}$ para todo $\mathbf{x} \in \mathbb{R}$, e $\int_{-\infty}^{\infty} \mathbf{f}_X(\mathbf{x})d\mathbf{x} = \mathbf{1}$ (ROSS, 2004; MAJOR, 2009).

Segundo Trosset (2009), variáveis aleatórias contínuas possuem uma função densidade de probabilidade. A FDP de uma variável aleatória contínua \mathbf{X} é dada por:

$$\mathbf{F}_X(\mathbf{x}) = P(X \leq \mathbf{x}) = \int_{-\infty}^{\mathbf{x}} \mathbf{f}_X(\mathbf{x})d\mathbf{x}, \quad -\infty < \mathbf{x} < \infty \quad (\text{B.8})$$

Montgomery e Runger (2010), e Brandimarte (2011), explicam que a média e variância de uma variável aleatória contínua são definidas similarmente às variáveis aleatórias discretas. A integração substitui o somatório nas definições. Assim, supondo que \mathbf{X} é uma variável aleatória contínua com uma FDP $\mathbf{f}_X(\mathbf{x})$, a **média** ou **valor esperado** de \mathbf{X} será:

$$\mu = E(X) = \int_{-\infty}^{\infty} \mathbf{x} \mathbf{f}_X(\mathbf{x})d\mathbf{x} \quad (\text{B.9})$$

A **variância** de \mathbf{X} será:

$$\begin{aligned} \sigma^2 = V(X) &= E[(X - \mu)^2] = \int_{-\infty}^{\infty} (\mathbf{x} - \mu)^2 \mathbf{f}_X(\mathbf{x})d\mathbf{x} \\ &= \int_{-\infty}^{\infty} \mathbf{x}^2 \mathbf{f}_X(\mathbf{x})d\mathbf{x} - \mu^2. \end{aligned} \quad (\text{B.10})$$

E o **desvio padrão** de \mathbf{X} será $\sigma = \sqrt{\sigma^2}$.

B.4 DISTRIBUIÇÃO NORMAL OU GAUSSIANA

O modelo amplamente utilizado para a distribuição de uma variável aleatória é a **distribuição normal**, a familiar “curva em forma de sino” como pode ser visto na Fig. (33). Sempre que um experimento aleatório é repetido, a variável aleatória que é igual ao resultado médio ao longo das repetições, tende a ter uma distribuição normal conforme o número de repetições cresce. Em 1733, De Moivre apresentou esse resultado fundamental conhecido como o **teorema do limite central**. Infelizmente, seu trabalho ficou perdido por um tempo e Gauss, de forma independente, desenvolveu a distribuição normal aproximadamente 100 anos depois. Apesar de mais tarde De Moivre ter sido creditado pela derivação, a distribuição normal é também referida como **distribuição Gaussiana** (MONTGOMERY; RUNGER, 2010).

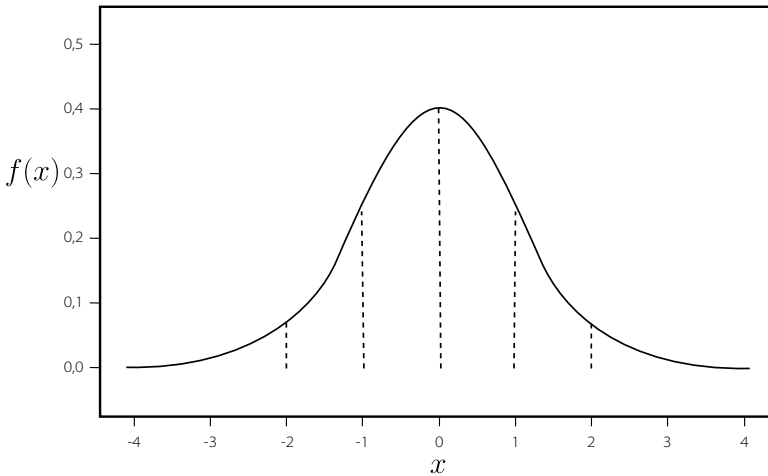


Figura 33 – Curva em forma de sino característica de uma distribuição normal. Baseado em Trosset (2009).

Segundo Trosset (2009) e Ross (2004), uma variável aleatória contínua \mathbf{X} é normalmente distribuída com média $-\infty < \mu < \infty$ e variância $\sigma^2 > 0$, denotada $\mathbf{X} \sim N(\mu, \sigma^2)$, se a FDP de \mathbf{X} for da forma:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \quad (\text{B.11})$$

A função distribuição de probabilidade de \mathbf{X} é dada por:

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(u-\mu)^2}{2\sigma^2}} du, \quad -\infty < x < \infty \quad (\text{B.12})$$

O valor esperado e a variância de \mathbf{X} são calculados por:

$$E[\mathbf{X}] = \mu \quad (\text{B.13})$$

$$\text{Var}(\mathbf{X}) = E[\mathbf{X}^2] - (E[\mathbf{X}])^2 = \sigma^2 \quad (\text{B.14})$$

A Eq. (B.11) não é tão extensivamente utilizada. Todavia, algumas propriedades bastante úteis sobre distribuição normal podem ser deduzidas diretamente da equação. A maior parte das seguintes propriedades podem ser observadas na Fig. (33).

1. $f_X(x) > 0$. Isto resulta que, para qualquer intervalo não-vazio (a, b) , $P(\mathbf{X} \in (a, b)) = \text{Area}_{(a,b)}(f_X) > 0$ e, portanto, $\mathbf{X}(S) = (-\infty, +\infty)$.
2. f_X é simétrico em relação a média μ . Por exemplo, $f_X(\mu+x) = f_X(\mu-x)$.
3. $f_X(x)$ diminui conforme $|x - \mu|$ aumenta.
4. $P(\mu - \sigma < \mathbf{X} < \mu + \sigma) \doteq 0,68$.
5. $P(\mu - 2\sigma < \mathbf{X} < \mu + 2\sigma) \doteq 0,95$.
6. $P(\mu - 3\sigma < \mathbf{X} < \mu + 3\sigma) \doteq 0,99$.

A área sob uma função densidade de probabilidade normal além de 3σ a partir da média é muito pequena. Pelo fato de mais de **99%** da probabilidade de uma distribuição normal estar entre o intervalo $(\mu - 3\sigma, \mu + 3\sigma)$, 6σ é frequentemente citado como a “largura” da distribuição normal (LARSON, 1982; MONTGOMERY; RUNGER, 2010).

Um fato importante sobre variáveis aleatórias normais é que se \mathbf{X} é normal com média μ e variância σ^2 , então $\mathbf{Y} = \alpha\mathbf{X} + \beta$ é normal com média $\alpha\mu + \beta$ e variância $\alpha^2\sigma^2$ (SOONG, 2004).

Segundo Ross (2004), uma variável aleatória normal \mathbf{X} com $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\sigma} = \mathbf{1}$ é chamada de **variável aleatória normal padrão**, denotada por \mathbf{Z} . Assim, se $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$, então:

$$\mathbf{Z} = \frac{(\mathbf{X} - \boldsymbol{\mu})}{\boldsymbol{\sigma}} \quad (\text{B.15})$$

Esta transformação é conhecida como **padronização**.

Soong (2004) explica que frequentemente faz-se necessário avaliar probabilidades associadas com uma variável aleatória normal $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$, tal como:

$$P(a < \mathbf{X} \leq b) = \frac{1}{\boldsymbol{\sigma}\sqrt{2\pi}} \int_a^b e^{-\frac{(x-\boldsymbol{\mu})^2}{2\boldsymbol{\sigma}^2}} dx \quad (\text{B.16})$$

Existem tabelas que permitem determinar facilmente probabilidades como a expressada pela Eq. (B.16).

No Apêndice A.1 é apresentado uma tabela com os valores para funções de distribuição de probabilidade para uma distribuição normal com $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\sigma} = \mathbf{1}$. A Tab. (A.1) fornece funções de distribuição de probabilidade $F_{\mathbf{Z}}(z)$ somente para os pontos onde $z \geq \mathbf{0}$. Segundo Soong (2004), os valores correspondentes à $z < \mathbf{0}$ são obtidos a partir da propriedade de simetria da distribuição normal padronizada através da relação:

$$F_{\mathbf{Z}}(-z) = 1 - F_{\mathbf{Z}}(z) \quad (\text{B.17})$$

A Tab. (A.1), a Eq. (B.15) e a Eq. (B.17) são suficientes para determinar probabilidades associadas com variáveis aleatórias normais com média e variância arbitrários.

Para concluir, um teorema útil, apresentado em (TROSSET, 2009), expõe a questão da soma de variáveis aleatórias normais independentes. Se $\mathbf{X}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\sigma}_1^2)$ e $\mathbf{X}_2 \sim N(\boldsymbol{\mu}_2, \boldsymbol{\sigma}_2^2)$ são independentes, então

$$\mathbf{X}_1 + \mathbf{X}_2 \sim N(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \boldsymbol{\sigma}_1^2 + \boldsymbol{\sigma}_2^2).$$